Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni

Definizione di un processo di anonimizzazione per la gestione di dati sanitari

Mario Sicuranza, Angelo Esposito, Mario Ciampi

RT-ICAR-NA-2017-04 **ottobre 2017**

Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR) – Sede di Napoli, Via P. Castellino 111, I-80131 Napoli, Tel: +39-0816139508, Fax: +39-0816139531, e-mail: napoli@icar.cnr.it, URL: *www.na.icar.cnr.it*

Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni

Definizione di un processo di anonimizzazione per la gestione di dati sanitari

Mario Sicuranza, Angelo Esposito, Mario Ciampi

Rapporto Tecnico N: **RT-ICAR-NA-2017-04** ***Data:*** **ottobre 2017**

Definizione di un processo di anonimizzazione per la gestione di dati sanitari

Mario Sicuranza, Angelo Esposito, Mario Ciampi

Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche
Via Pietro Castellino, 111 – 80131 Napoli, Italia
E-mail: {mario.sicuranza, angelo.esposito, mario.ciampi}@icar.cnr.it

Abstract

L'elaborazione di enormi quantità di informazioni generate e raccolte mediante sistemi di ICT per la sanità offre grosse opportunità nella possibilità di generare nuova conoscenza, da utilizzare in applicazione software specifiche, ad esempio applicazioni per il miglioramento dei percorsi diagnostico-terapeutici assistenziali per malattie croniche. Nel contesto dell'elaborazione delle informazioni sanitarie, un aspetto fondamentale da tenere in conto sin dalle prime fasi della progettazione dei sistemi informativi riguarda la tutela del diritto di privacy di ogni cittadino/paziente. Per preservare tale diritto è possibile, a monte del trattamento dei dati, utilizzare tecniche di anonimizzazione. Tuttavia, nel contesto sanitario, non basta l'applicazione di tecniche generiche di anonimizzazione ma è necessario definire un approccio specifico, che, in funzione al contesto di utilizzo del dato sanitario, permetta di trattare in maniera differente l'informazione e di applicare tecniche di anonimizzazione specifiche. In questo rapporto tecnico è descritto un processo di consultazione dei dati sanitari, gestiti mediante sistemi informativi, in forma anonimizzata mediante l'applicazione di un algoritmo che permette in maniera flessibile e precisa l'anonimizzazione dei dati sanitari richiesti nei diversi contesti di elaborazione. L'applicazione del processo descritto e dell'algoritmo definito consente il recupero di documenti sanitari in forma anonimizzata, garantendo in questo modo la privacy del paziente a cui si riferiscono tali documenti. In particolare, l'algoritmo di anonimizzazione sviluppato realizza una de-identificazione dei dati personali applicando la tecnica di L-L-diversità a più livelli e discriminando le tipologie di informazioni che sono gestite. Il processo e l'algoritmo sono stati sperimentati nell'ambito del progetto di ricerca "eHealthNet" e la loro applicazione ha consentito di eliminare le associazioni tra uno specifico paziente e le informazioni cliniche presenti in diverse tipologie di documenti clinici.

Keywords: Processo sanitario, Anonimizzazione, Privacy, Information Retrieval.

1. Introduzione

L'applicazione dell'ICT nel contesto sanitario offre numerose possibilità, soprattutto per quanto riguarda l'utilizzo di informazioni sanitarie per finalità di cura, ricerca e governo. Nell'ambito dell'elaborazione delle informazioni è di fondamentale importanza garantire la privacy al cittadino/paziente a cui si riferiscono le informazioni sanitarie trattate. Lo scopo per il quale vengono richieste le informazioni determina anche il tipo di elaborazione e meccanismo da adottare al fine di tutelare la privacy, la finalità può essere ad esempio quella di ricerca. Per la finalità di cura esistono diversi meccanismi che possono essere utilizzati per regolare l'accesso alle informazioni unicamente ai soggetti autorizzati, come ad esempio i modelli di controllo degli accessi [9,14] che possono essere utilizzati per trattare informazioni prelevate da sistemi per il monitoraggio di pazienti [18]. Esistono inoltre sistemi che utilizzano dati sanitari in piena sicurezza, permettendo la sottoscrizione ai medici autorizzati ai dati di interesse [4]. Tutti questi sistemi, così come quello mostrato in [5,20] consentono di condividere in piena sicurezza i dati mediante una federazione di sistemi informativi, tali informazioni possono essere condivise in piena sicurezza anche mediante sistemi basati sulle notifiche, come ad esempio in [21].

Per la finalità di governo e di ricerca esistono tecniche differenti, che consentono di eliminare l'associazione tra il paziente e i dati sanitari che lo riguardano. L'anonimizzazione [1] rappresenta una delle tecniche maggiormente utilizzate atte a garantire privacy nella elaborazione delle informazioni. L'anonimizzazione realizza una de-identificazione dei dati personali e può essere utilizzata per i documenti clinici allo scopo di eliminare l'associazione tra uno specifico paziente e le informazioni cliniche presenti in esso. In questo rapporto tecnico è descritto un processo per la consultazione dei dati anonimizzati, che risulta facilmente implementabile e integrabile in una architettura a servizi (SOA) [2].

Il capitolo 2 descrive mediante lo standard BPMN il processo definito, il quale utilizza un algoritmo specifico di anonimizzazione, che permette in maniera flessibile e precisa l'anonimizzazione dei dati sanitari. L'algoritmo progettato e sviluppato, descritto nel capitolo 3, combina diverse tecniche note in letteratura in maniera specifica, in funzione della tipologia di documenti richiesti e alla motivazione di accesso.

Il capitolo 4 è dedicato alla descrizione di una possibile interfaccia programmatica, mentre il capitolo 5 è dedicato alla descrizione di una componete software che utilizza le interfacce programmatiche. Il documento infine definisce una serie di test per la componente, che permettono la validazione della stessa in conformità al processo definito.

2. Definizione del processo di Consultazione dei dati Anonimizzati mediante BPMN 2.0

Questo capitolo mostra, mediante una rappresentazione grafica il processo di consultazione dei dati anonimizzati. La rappresentazione grafica è realizzata mediante lo standard BPMN 2.0 [3].
Il processo può facilmente essere implementato come servizio in una architettura SOA, come ad esempio quella descritta in [13].

2.1. Descrizione del processo

Il processo di **Consultazione Dati Anonimizzati** permette ad un Utente autenticato, tramite il sistema Client, di richiedere al sistema informativo sanitario dei dati anonimizzati che soddisfano determinati parametri di ricerca. Il sistema renderà fruibili in maniera anonima tali dati all'utente che ne ha fatto la richiesta.

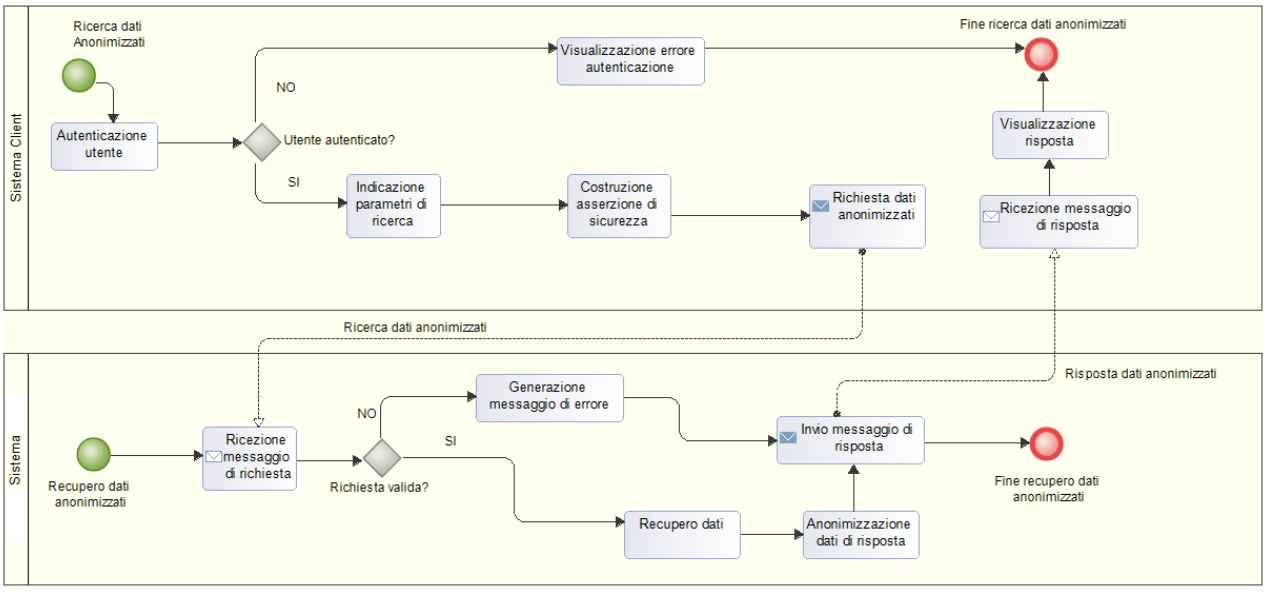

Figura 1: BPMN del processo Consultazione Dati Anonimizzati

Il processo di business è rappresentato e costituito da differenti attività, che sono descritte di seguito.

- **Autenticazione utente,** il nodo che realizza questa attività è il **Sistema Client.** L'attività prevede l'autenticazione dell'utente sul Sistema Client.
- **Visualizza errore autenticazione,** il nodo che realizza questa attività è il **Sistema Client.** L'attività visualizza un messaggio di errore per l'autenticazione fallita da parte dell'utente. Questa attività viene eseguita solo nel caso di autenticazione fallita da parte dell'utente.
- **Indicazione parametri di ricerca,** il nodo che realizza questa attività è il **Sistema Client.** L'attività prevede l'indicazione dei parametri di ricerca dei documenti indicati da parte dell'utente.
- **Costruzione asserzione di sicurezza,** il nodo che realizza questa attività è il **Sistema Client**. L'attività prevede la costruzione di una asserzione di attributo che il sistema client utilizza nel messaggio di richiesta.
- **Richiesta dati anonimizzati,** il nodo che realizza questa attività è il **Sistema Client.** L'attività prevede l'invio della richiesta dei dati anonimizzati che rispettano i criteri di ricerca indicati nella attività precedente.
- **Ricezione messaggio di richiesta,** il nodo che realizza questa attività è il **Sistema Informativo.** L'attività prevede la ricezione del messaggio di richiesta dei dati anonimizzati da parte del **Sistema Client**.
- **Generazione messaggio di errore,** il nodo che realizza questa attività è il **Sistema Informativo.** L'attività prevede la generazione di un messaggio di errore nel caso la richiesta non sia formulata in maniere corretta.
- **Recupero dati,** il nodo che realizza questa attività è il **Sistema Informativo.** L'attività prevede il recupero dei dati che soddisfano i criteri di ricerca indicati nella richiesta inviata dal **Sistema Client**.
- **Anonimizzazione dati di risposta,** il nodo che realizza questa attività è il **Sistema Informativo.** L'attività prevede l'anonimizzazione dei dati di risposta da inviare al **Sistema Client** mediante l'algoritmo di anonimizzazione descritto di seguito nel presente documento.
- **Invio messaggio di risposta,** il nodo che realizza questa attività è il **Sistema Informativo.** L'attività prevede l'invio del messaggio di risposta alla richiesta effettuata dal **Sistema Client**. La risposta può contenere un messaggio di errore oppure i dati anonimizzati richiesti.
- **Ricezione messaggio di risposta,** il nodo che realizza questa attività è il **Sistema Client.** L'attività prevede la ricezione del messaggio di risposta inviata dal **Sistema Informativo**.
- **Visualizza messaggio di risposta,** il nodo che realizza questa attività è il **Sistema Client.** L'attività prevede la visualizzazione dei dati anonimizzati ottenuti come risposta alla richiesta effettuata dal **Sistema Client** al **Sistema Informativo** oppure un messaggio di errore.

3. Algoritmo di Anonimizzazione

3.1. Stato dell'arte

L'elaborazione automatica di dati sanitari raccolti mediante sistemi informativi offre vantaggi alla società in termini economici e sociali. La combinazione di dati può generare nuova conoscenza, che può portare ad applicazioni non ancora valutate. L'elaborazione delle informazioni sanitarie è realizzabile solo se è rispettato il diritto di ogni cittadino alla protezione dei propri dati personali e della propria sfera privata. Ad oggi, l'anonimizzazione rappresenta uno degli strumenti maggiormente utilizzati per l'elaborazione delle informazioni garantendo ad ogni cittadino la propria privacy. Il concetto di privacy è largamente utilizzato in diversi campi, ed ha differenti significati a seconda della disciplina in cui viene utilizzato. Un articolo del 1890 "The Right to Privacy", definisce la privacy di un individuo come il diritto di essere lasciato solo, questa rappresenta la prima pubblicazione che evoca il diritto di privacy [12]. Una prima tipologia di tecniche che permettono di regolare l'accesso ai dati, sono i cosiddetti meccanismi di controllo degli accessi. Esistono diversi modelli di controllo degli accessi, alcuni dei quali forniscono direttamente al paziente cui si riferiscono i documenti clinici la possibilità di definire regole per l'accesso ai propri dati sanitari [7,9].

Nella valutazione delle tecniche e degli algoritmi noti in letteratura, e di conseguenza nella definizione di una specifica tecnica di anonimizzazione, sono stati analizzati i seguenti aspetti critici:

- **L'individuazione (**della persona**)**, ovvero la possibilità di estrapolare alcuni (o tutti i) dati che consentono l'identificazione di una persona;
- **La correlazione**, ovvero la possibilità di collegare diversi dati relativi ad una stessa persona;
- **La deduzione**, ovvero la possibilità di evincere il valore di un dato a partire da altre informazioni correlate.

Nella fase di definizione dell'algoritmo di anonimizzazione è stata svolta una analisi del modello dei dati sanitari da elaborare, le diverse tipologie di dati sono state correlate con diversi livelli di sicurezza e gradi di anonimizzazione. Nei prossimi paragrafi sono descritti brevemente gli algoritmi e le tecniche di anonimizzazione che si prestano bene a poter essere utilizzati nel contesto sanitario italiano. L'algoritmo di anonimizzazione sviluppato è mostrato nel capitolo successivo.

Di seguito sono descritte brevemente le principali tecniche note in letteratura utilizzate per realizzare l'anonimizzazione delle informazioni personali, che possono essere facilmente applicabili al contesto dell'e-health.

- **Data masking** [6]: rappresenta una tecnica di anonimizzazione che prevede la cancellazione dei principali identificativi personali, come ad esempio il nome, la data di nascita, questa tecnica è utilizzata ad esempio nelle basi di dati giuridiche;
- **Randomizzazione** [15]**:** rappresenta una tecnica di anonimizzazione che ha lo scopo di modificare i dati al fine di eliminare la relazione tra i dati e la persona. Esistono numerose metodologie di anonimizzazione che utilizzano questa tecnica a diversi livelli nella struttura dell'informazione.
- **Pseudoanimizzazione** [16]**:** permette di sostituire un valore di un attributo con un altro valore, in questo modo si ottiene una pseudoanonimizzazione, e non una completa anonimizzazione, in quanto con l'opportuna tabella di sostituzione dei valori è possibile ottenere le informazioni originali. La pseudoanonimizzazione si può definire come la tecnica con la quale si sostituisce un attributo univoco di un dato con un altro. La persona potrebbe, comunque, essere identificata in maniera indiretta;
- **Generalizzazione** [8]**:** rappresenta una tecnica di anonimizzazione che ha l'obiettivo di generalizzare gli attributi associati alle persone. Ad esempio, l'informazione relativa ad una data di nascita può

essere generalizzata utilizzando unicamente l'anno di nascita ed evitando l'indicazione del giorno e del mese.

Inoltre, esistono diverse tecniche evolute per l'anonimizzazione [10,11] che esulano dall'obiettivo di questo rapporto tecnico.

La tecnica scelta per la realizzazione dell'algoritmo utilizza diverse metodologie proprie della generalizzazione. Per questo motivo, di seguito sono descritte tecniche di generalizzazione che sono state valutate per l'impiego nell'algoritmo di anonimizzazione definito.

- **Aggregazione e k-anonimato** [17]**:** queste tecniche, tramite l'aggregazione con k persone diverse, tentano di impedire l'individuazione di una persona. Facendo condividere lo stesso valore con k persone risulta più difficile l'individuazione di una specifica persona. La generalizzazione quindi permette di far condividere un dato valore di un attributo per un numero di persone maggiori.

  Il difetto principale del modello del k-anonimato è che non protegge da attacchi deduttivi. Inoltre, con l'intersezione di diversi gruppi rappresentati da diversi attributi può essere ancora più semplice individuare la persona.

- **L-L-diversità** [19]**:** la L-L-diversità estende la tecnica di tipo k-anonimato per rendere poco efficaci gli attacchi tramite deduzione deterministica facendo sì che in ciascuna classe di equivalenza ci siano almeno L diversi valori di L attributi. La L-L-diversità è soggetta ad attacchi tramite deduzione probabilistica.

- **T-Vicinanza:** questa tecnica rappresenta una evoluzione della **L-L-diversità**, in quanto l'obiettivo è quello di creare classi equivalenti che siano simili agli attributi iniziali. Utile quando si vuole che i valori ottenuti siano quanto più vicini a quelli di partenza. Questa tecnica impone che non solo devono esistere almeno L valori diversi all'interno di ogni classe di equivalenza, così come indicato dalla tecnica di **L-L-diversità**, ma anche che ogni valore è rappresentato tante volte quante sono necessarie per rispecchiare la distribuzione iniziale di ciascun attributo.

Nella maggior parte dei casi non è possibile fornire raccomandazioni minime circa i parametri da utilizzare, in quanto per ogni modello dei dati è necessaria una specifica analisi. L'algoritmo sviluppato e descritto nel prossimo paragrafo tiene in conto del modello dei dati sanitari da elaborare, correlando e combinando alcune delle tecniche e degli algoritmi descritti sopra.

3.2. Algoritmo di Anonimizzazione

L'algoritmo di anonimizzazione progettato e sviluppato prevede la combinazione e l'uso delle tecniche, note in letteratura, brevemente descritte nel paragrafo precedente.

L'algoritmo definito si comporta in modo differente in funzione del tipo di attributo da trattare applicando la tecnica di anonimizzazione più adatta a seconda della tipologia di dato da trattare.

Nella Figura 2 è descritto l'algoritmo mediante un flow chart, che ne fornisce una rappresentazione grafica. Considerando il documento sanitario da anonimizzare, l'algoritmo analizza, per ogni attributo del documento, la tipologia dell'attributo e verifica la presenza dello stesso nella tabella di trasformazione. Se la tipologia dell'attributo da trattare è presente nella tabella, l'algoritmo applica la trasformazione indicata e sostituisce il dato trasformato nel documento sanitario. Se la tipologia di attributo da trattare non è presente nella tabella, l'algoritmo non modifica l'attributo, in quanto non appartiene a nessuna delle tipologie di dato da anonimizzare.

Di seguito è mostrato un esempio di tabella di trasformazione, in cui si evince la correlazione tra la tipologia di attributo e la tecnica di anonimizzazione da applicare.

Tipologia di attributo	Tecnica di anonimizzazione da applicare
Indirizzi	Applicazione dell'algoritmo di LL-diversità
Date	Applicazione dell'algoritmo di LL-diversità
Identificativi	Cancellazione del valore dell'attributo

Tabella 1 - Tabella di Trasformazione (Esempio)

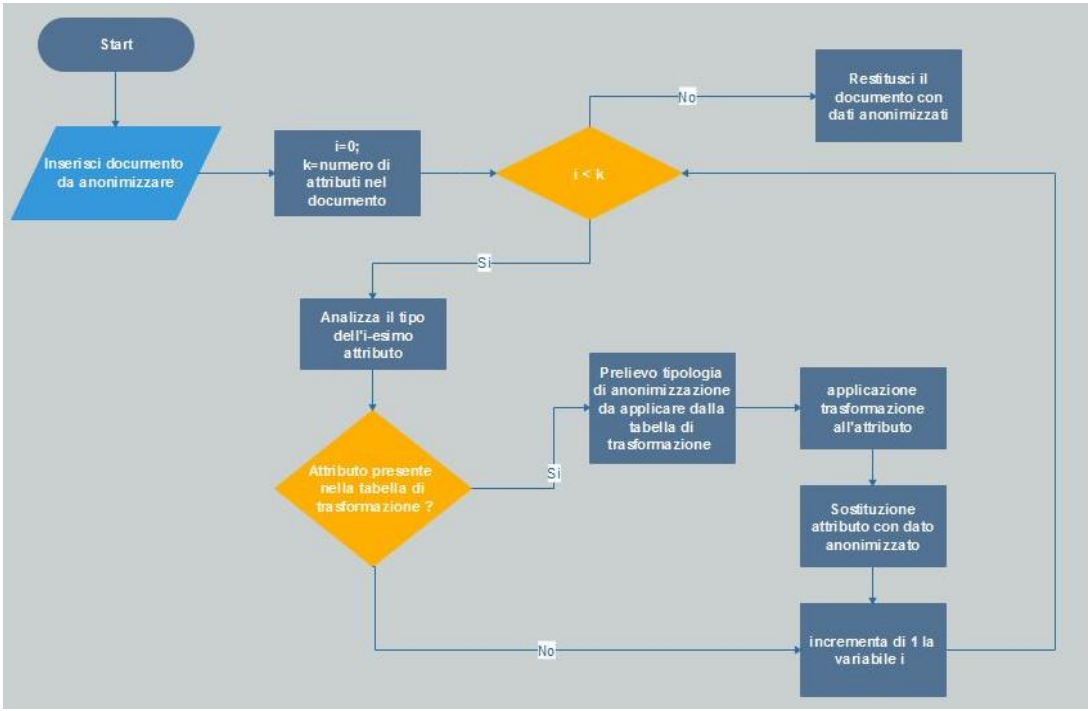

Figura 2 - Flow chart algoritmo di Anonimizzazione

La particolarità dell'algoritmo di anonimizzazione, presentato in questo paragrafo, consiste nell'utilizzo di una tabella di trasformazione che rende l'algoritmo molto flessibile e utilizzabile nei diversi contesti operativi. Infatti, la tabella consente di applicare diverse tecniche di anonimizzazione a diverse tipologie di attributo. Nell'esempio in Tabella 1, per quanto riguarda le date indicati nel documento, si applica un algoritmo della classe LL-diversità. Alcune applicazioni reali dell'algoritmo sono schematizzate di seguito.

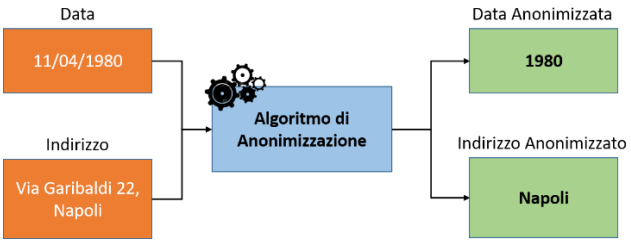

Figura 3 - Applicazione dell'Algoritmo di Anonimizzazione

4. Interfaccia Consultazione Dati Anonimizzati

In questo capitolo è presentata una interfaccia programmatica attraverso la quale è possibile usufruire dei servizi di ricerca e recupero dati/documenti anonimizzati. La Tabella 2 mostra la descrizione dell'interfaccia associata al servizio che permette di realizzare il processo mostrato al capitolo 2, specificando le precondizioni,

i parametri di input e output e le eventuali eccezioni.
Questa Interfaccia è costituita da due distinte operazioni, l'operazione di Ricerca Dati Anonimizzati e l'operazione di Recupero Dati Anonimizzati.

4.1.1. Ricerca Dati Anonimizzati

Campo	Descrizione
Descrizione	Effettua l'operazione di ricerca dati sanitari in forma anonimizzata sulla base di un insieme di parametri di ricerca che l'operatore può indicare per individuare informazioni sanitarie di interesse.
Precondizioni	L'utente deve essere correttamente autenticato.
Input	Parametri di ricerca.
Output	Metadati in forma anonimizzata per il recupero dei documenti sanitari anonimizzati.
Post-Condizioni	L'utente riceve i metadati relativi ai documenti sanitari che corrispondono ai criteri di ricerca indicati.
Eccezioni	• Accesso negato (ad esempio diritti utente non sufficienti) • Asserzioni non valide o parametri non coerenti.

Tabella 2 - Descrizione Operazione Ricerca Dati Anonimizzati

Messaggio di richiesta

Nella tabella di seguito i campi presenti nel messaggio di richiesta con una breve descrizione.

Campo	Descrizione	Obbligatorietà
Lista dei parametri di ricerca	Rappresenta l'elenco dei parametri da utilizzare per effettuare la ricerca.	Si

Tabella 3 - Parametri, messaggio di richiesta Operazione Ricerca Dati Anonimizzati

Messaggio di risposta

Nella tabella di seguito i campi presenti nel messaggio di risposta con una breve descrizione in caso di risposta con successo.

Campo	Tipologia	Descrizione
Lista dei metadati in forma anonimizzata	Stringa	Lista di metadati relativi ai documenti individuati in funzione dei criteri di ricerca indicati nel messaggio di richiesta.

Tabella 4 - Parametri, messaggio di risposta con successo Operazione Ricerca Dati Anonimizzati

Messaggio di risposta di fallimento

Nella tabella di seguito i campi presenti nel messaggio di risposta con una breve descrizione in caso di risposta con fallimento.

Campo	Descrizione	Obbligatorietà
Stato risposta	Successo/Fallimento	Si
Codice errore	Sono codici interni al sistema	Si

Tabella 5 - Parametri, messaggio di risposta con fallimento Ricerca Dati Anonimizzati

4.1.2. Recupero Dati Anonimizzati

Campo	Descrizione
Descrizione	Effettua l'operazione di recupero dei dati sanitari in forma anonimizzata tramite un insieme di identificativi dei documenti sanitari.
Precondizioni	• L'utente deve essere correttamente autenticato. • L'utente deve conoscere l'insieme degli identificativi dei documenti che intende recuperare in forma anonimizzata.
Input	• Identificativi dei documenti da recuperare. • Parametro che specifica la modalità di recupero (separato o aggregato)
Output	Uno o più documenti aggregati in forma anonimizzata.
Post-Condizioni	L'utente riceve l'insieme dei documenti in forma singola o aggregata identificati mediante l'insieme degli identificativi indicati.
Eccezioni	• Permesso di accesso negato (diritti utente non sufficienti) • Asserzioni o riferimenti ai documenti non validi.

Tabella 6 - Descrizione operazione Recupero dati anonimizzati

Di seguito si mostrano gli elementi presenti nei messaggi di richiesta e risposta per l'operazione Recupero dati anonimizzati.

Messaggio di richiesta

Nella tabella di seguito i campi presenti nel messaggio di richiesta con una breve descrizione.

Campo	Tipologia	Descrizione	Obbligatorietà
Lista degli identificativi dei documenti	Stringhe	È una stringa, che specifica la lista degli identificativi dei documenti da recuperare in forma anonimizzata.	Si
Opzione aggregazione	Booleano	Indica, se a true, che i dati devono essere estratti dai documenti identificati dai parametri indicati nella	No

		“lista degli identificativi” e aggregati in un unico documento. Se questo parametro è a false indica che i documenti devono essere anonimizzati e restituiti dalla piattaforma nella loro formato originale.	

Tabella 7- Parametri, messaggio di richiesta Operazione Recupero dati anonimizzati

Messaggio di risposta con successo

Nella tabella di seguito i campi presenti nel messaggio di risposta, in caso di successo, con una breve descrizione.

Campo	Tipologia	Descrizione
Insieme dei documenti anonimizzati	Documenti	Insieme di documenti anonimizzati richiesti. Questi documenti possono essere a discrezione dell’utente in forma aggregata o di singolo documento.

Tabella 8 - Parametri, messaggio di risposta Operazione Recupero dati anonimizzati

Messaggio di risposta di fallimento

Nella tabella di seguito i campi presenti nel messaggio di risposta, in caso di fallimento, con una breve descrizione.

Campo	Tipologia	Obbligatorietà
Codice errore	specifico per messaggio	Si

Tabella 9 - Parametri, messaggio di risposta con fallimento Operazione Recupero dati anonimizzati

5. Componente software per la consultazione anonima dei documenti

In questo capitolo è fornita una descrizione di una possibile componente software progettata per la consultazione anonima dei documenti che utilizza il servizio di ricerca e recupero dati/documenti anonimizzati. Questa componente installata presso un sistema informativo è capace di offrire il servizio di Consultazione anonima dei Documenti sanitari. La componente software utilizza le componenti di Registry e di Repository, che possono essere sia installate presso un sistema interno sia installate presso un sistema esterno.

5.1. Iterazioni e funzionalità nel processo di Consultazione Anonima Documenti

Il sequence diagram mostrato in Figura 4 rappresenta graficamente le interazioni tra i vari attori che partecipano al processo di ricerca e consultazione dati anonimizzati.

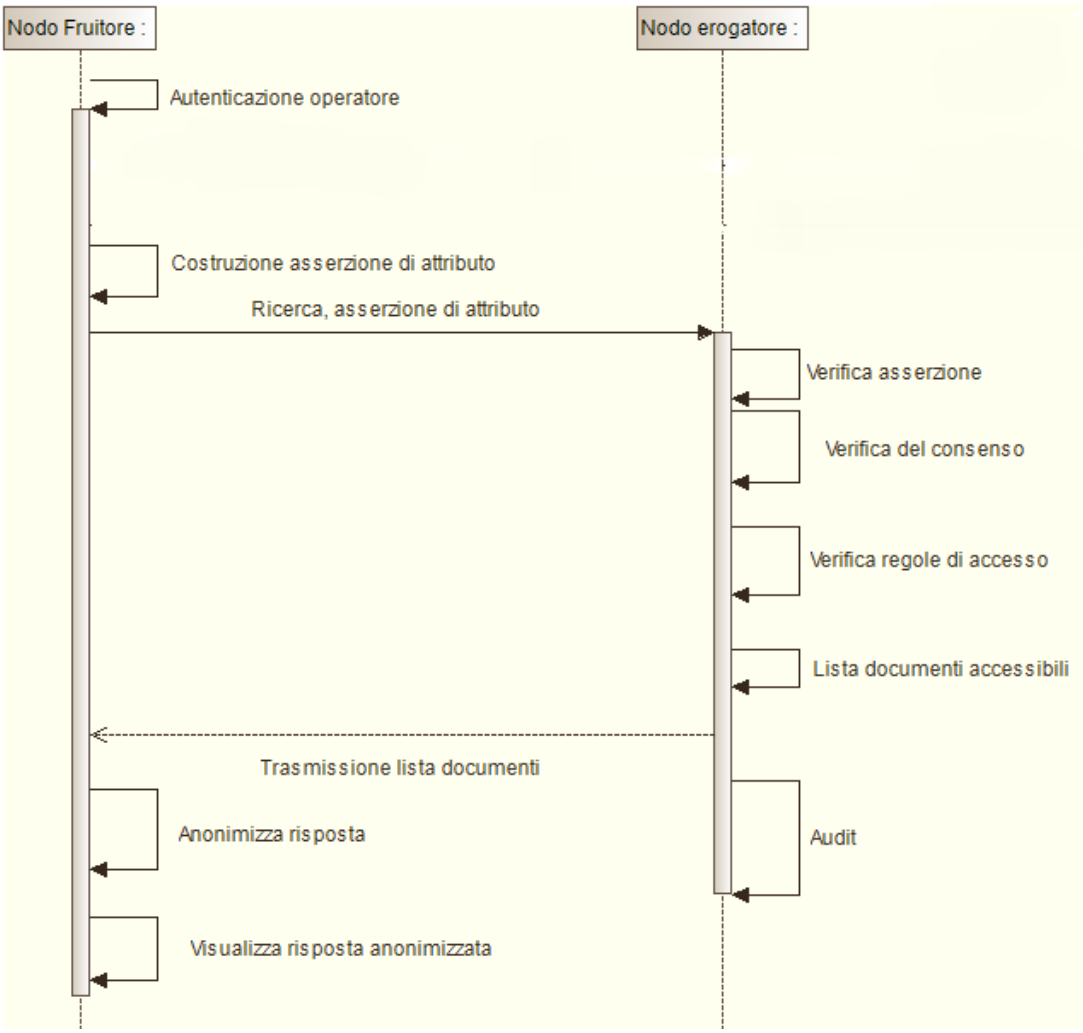

Figura 4: Consultazione Anonima Documenti

Di seguito la tabella delle componenti.

Componente	Descrizione
Nodo Fruitore	Usufruire del servizio messo a disposizione del nodo erogatore L'anonimizzazione è realizzata dal nodo fruitore, che provvede, prima di inviare la risposta al client che ha effettuato la richiesta di anonimizzare la risposta.
Nodo Erogatore	Gestisce documenti e dati sanitari. Ha la responsabilità di verificare il possesso dei diritti di accesso ai documenti sanitari da parte del richiedente.

Tabella 10: Componenti per la Consultazione Anonima dei Documenti e Dati

5.2. Tecnologie per lo sviluppo software e campagna di test

Le tecnologie utilizzate per l'implementazione del servizio e della componente software sono sintetizzate nella seguente tabella.

Tecnologia	Versione	Descrizione
Java API for XML Web	2.0	JAX-WS è insieme di procedure del linguaggio di programmazione

Services		Java dedicate allo sviluppo di servizi web
Security Assertion Markup Language (SAML)	2.0	standard informatico per lo scambio di dati tramite asserzioni tra domini di sicurezza distinti.
PostgreSQL	9.5	PostgreSQL è un DBMS ad oggetti, utilizzato per la gestione dei repository e registry.

5.3. Campagna di test

In questo paragrafo sono descritti i casi di test definiti per la valutazione del corretto funzionamento del servizio "Consultazione Anonima Documenti" implementato ed inserito nella piattaforma di servizi definita nell'ambito del progetto "eHealthNet".

N.	Descrizione del test	Prerequisito	Input	Risultato atteso
C1	Consultazione anonima dei documenti con successo	Parametri di ricerca validi e noti	Insieme di Parametri di ricerca validi	La consultazione anonima va a buon fine, viene restituito un messaggio contenente la lista di metadati anonimizzati associati ai documenti che soddisfano i criteri di ricerca
C2	Richiesta di consultazione anonima non riconosciuta	-	Almeno un parametro di ricerca non noto	La consultazione anonima non va a buon fine, viene restituito un messaggio contenente l'indicazione dell'errore riscontrato.
C3	Richiesta di consultazione anonima non riconosciuta	-	Almeno un parametro di ricerca non valido	La consultazione anonima non va a buon fine, viene restituito un messaggio contenente l'indicazione dell'errore riscontrato.

6. Conclusioni

In questo rapporto tecnico è stato presentato un processo di consultazione dei dati sanitari, gestiti mediante sistemi informativi sanitari, in forma anonimizzata mediante l'applicazione di une un algoritmo che permette in maniera flessibile e precisa l'anonimizzazione dei dati sanitari richiesti nei diversi contesti di elaborazione. L'applicazione del processo descritto e l'algoritmo definito consente il recupero di documenti sanitari in forma anonimizzata, garantendo la privacy del paziente a cui si riferiscono tali documenti. In particolare, l'algoritmo di anonimizzazione sviluppato realizza una de-identificazione dei dati personali applicando diverse tecniche tra cui la tecnica di LL-diversità a diversi livelli.
Il processo di "Consultazione dei dati Anonimizzati" è stato realizzato mediante la progettazione e l'implementazione di un servizio web, le cui interfacce programmatiche sono descritte in questo rapporto tecnico. In particolare, tali interfacce contemplano le operazioni di ricerca e recupero dati anonimizzati, grazie alle quali è possibile rispettivamente ricercare dati anonimizzati su una popolazione di pazienti e recuperare le informazioni in forma anonimizzata applicando l'algoritmo di anonimizzazione descritto in questo rapporto tecnico.
Il processo e l'algoritmo sono stati sperimentati nell'ambito del progetto di ricerca "eHealthNet" e la loro applicazione ha consentito di eliminare le associazioni tra uno specifico paziente e le informazioni cliniche presenti in diverse tipologie di documenti clinici in formato HL7 CDA 2.

Riferimenti

1. Anonymisation: managing data protection risk code of practice, disponibile al seguente link: https://ronchilegal.eu/wp-content/uploads/2017/12/anonymisation-code-ICO.pdf
2. W3C - Web Open Standards, Service Oriented Architecture (SOA), disponibile al seguente link: https://www.w3.org/standards/
3. About the business process model and notation specification version 2.0 BPMN 2.0, OMG standards, disponibile al seguente link: https://www.omg.org/spec/BPMN/2.0/
4. Esposito, C., Ciampi, M., De Pietro, G., Donzelli, P. (2012, July). Notifying medical data in health information systems. In Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems (pp. 373-374).
5. Ciampi, M., De Pietro, G., Esposito, C., Sicuranza, M., Mori, P., Gebrehiwot, A., Donzelli, P. (2012). On securing communications among federated health information systems. In proceedings of the 31st International Conference on Computer Safety, Reliability and Security, SAFECOMP 2012 Workshops, Volume 7613, pp. 235-246.
6. "Data Masking vs. Data Encryption". www.iri.com. Innovative Routines International. Retrieved 24 August 2017.
7. Sicuranza, M., Esposito, A., Ciampi, M. (2015). An access control model to minimize the data exchange in the information retrieval. Journal of Ambient Intelligence and Humanized Computing, 6(6), 741-752.
8. Big Data e Privacy by design, Big data e privacy by design. Anonimizzazione, pseudonimizzazione, sicurezza, Maurizio Naldi, Giuseppe D'Acquisto. 2017
9. Sicuranza, M., Ciampi, M. (2014, November). A semantic access control for easy management of the privacy for EHR systems. In P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2014 Ninth International Conference on (pp. 400-405). IEEE.
10. Xiao Pan, Jianliang Xu, Xiaofeng Meng: Protecting Location Privacy against Location-Dependent Attacks in Mobile Services. IEEE Trans. Knowl. Data Eng. 24(8): 1506-1519 (2012)
11. Xiao Pan, Lei Wu, Chunhui Piao, et al: P3RN: Personalized Privacy Protection using Query Semantics over Road Networks. In procedings of Waim 2014, 2014
12. Warren S, Brandeis L. The right to privacy[J].Harvard law review,1890,4(5):193-220
13. Chiaravalloti, M. T., Ciampi, M., Pasceri, E., Sicuranza, M., De Pietro, G., Guarasci, R. (2015, January). A model for realizing interoperable EHR systems in Italy. In Proceedings of the 15th International HL7 Interoperability Conference (pp. 13-22).
14. Sicuranza, M., Esposito, A., Ciampi, M. (2014). A patient privacy centric access control model for EHR systems. International Journal of Internet Technology and Secured Transactions, 5(2), 163-189.
15. Kang, Minsoo, Brian G Ragan, and Jae-Hyeon Park. “Issues in Outcomes Research: An Overview of Randomization Techniques for Clinical Trials.” Journal of Athletic Training 43.2 (2008): 215–221. Print.
16. Bolognini, C. Bistolfi, Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation, Computer Law & Security review, 2016
17. N. Li, T. Li and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," 2007 IEEE 23rd International Conference on Data Engineering, Istanbul, 2007, pp. 106-115. doi: 10.1109/ICDE.2007.367856
18. A Coronato, G De Pietro, Situation awareness in applications of ambient assisted living for cognitive impaired people, Mobile Networks and Applications 18 (3), 444-453
19. A. Machanavajjhala, J. Gehrke, D. Kifer and M. Venkitasubramaniam, "L-diversity: privacy beyond k-anonymity," 22nd International Conference on Data Engineering, Atlanta, GA, USA, 2006, pp. 24-24. doi: 10.1109/ICDE.2006.1
20. F Amato, G De Pietro, M Esposito, N Mazzocca, An integrated framework for securing semi-structured health records, Knowledge-Based Systems, 79, 99-117
21. C Esposito, M Ciampi, G De Pietro, An event-based notification approach for the delivery of patient medical information, Information Systems, 39, 22-44