Contents

[Introduzione 2](#_Toc466769972)

[1 MyData, Big Data 3](#_Toc466769973)

[2 Smart Mobility 4](#_Toc466769974)

[2.1 Mobility as a Service 4](#_Toc466769975)

[2.1.1 Mobility Profile e Journey Planner 5](#_Toc466769976)

[2.2 Smart Mobility for All 5](#_Toc466769977)

[2 ½ GDPR?? 6](#_Toc466769978)

[3 Architettura MyData 6](#_Toc466769979)

[3.1 Entità fondamentali 6](#_Toc466769980)

[3.2 Service Registry, Service Linking 7](#_Toc466769981)

[3.2.1 Oauth 2.0 7](#_Toc466769982)

[3.3 Autorizzazioni e Consent 7](#_Toc466769983)

[3.3.1 Kantara Consent & Information Sharing Work Group 8](#_Toc466769984)

[3.3.2 UMA 8](#_Toc466769985)

[3.4 Personal Data Storage 9](#_Toc466769986)

[4 Analisi e Design 10](#_Toc466769987)

[4.1 Accounting E Servizi 10](#_Toc466769988)

[4.2 Operatore MyData 11](#_Toc466769989)

[4.3 Service Registry 11](#_Toc466769990)

# Introduzione

Negli ultimi anni, con la diffusione degli smartphone e l’aumento dell’accesso a Internet, la tecnologia è diventata sempre più legata all’utente dietro lo schermo, con esempi che possono spaziare dai social network ai dispositivi di monitoraggio della salute e dell’attività fisica. La quantità di informazioni raccolte tramite questi servizi non si limita a ciò che l’utente sceglie di condividere (foto, email, transazioni online), ma comprende anche dati osservati (abitudini di navigazione, dati di geolocalizzazione) e deduzioni (*targeted advertising*, previsioni sul flusso del traffico).

L’importanza dei dati personali è diventata tale da richiedere l’esistenza di leggi che ne regolamentassero l’utilizzo, come il regolamento generale sulla protezione dei dati (GDPR, *General Data Protection Regulation- Regolamento UE 2016/679)*.

In questo contesto si sviluppa il modello di *MyData*, fondato sull’idea secondo cui ogni utente ha il controllo sui propri dati personali, ed è a conoscenza dell’utilizzo che ne viene fatto, in modo trasparente. D’altra parte, l’architettura *MyData* vuole offrire al mercato e alle imprese un contesto di sviluppo di applicazioni software in cui vengono rispettate le leggi in materia di protezione dei dati sensibili, favorendo l’interoperabilità fra di esse.

L’approccio *MyData*, seppur innovativo, non è senza precedenti. Molti dei suoi concetti fondamentali sono anche alla base dei *personal cloud*, come ad esempio la possibilità di contenere in modo sicuro i dati personali dell’utente e offrire interoperabilità fra diversi servizi, utilizzando concretamente i dati memorizzati.

Questa tesi si propone di studiare i concetti fondamentali alla base del modello di *MyData*, e di implementare un prototipo di gestore di dati personali che ne rispetti le specifiche. A titolo di esempio, si è scelto di utilizzare come servizio “consumatore” di dati personali un sistema di previsione del viaggio più probabile.

L’obiettivo è quello di realizzare un sistema che permetta all’utente di monitorare l’utilizzo dei suoi dati personali in tempo reale ed in modo trasparente, attraverso l’implementazione di una politica di controllo degli accessi. Inoltre, il gestore opera in un contesto in cui il rispetto di politiche di sicurezza è fondamentale: anche se la realizzazione di un sistema sicuro non rientra negli obiettivi del lavoro di tesi, sono presenti alcuni accorgimenti che puntano in questa direzione.

# 1 MyData, Big Data

Il progetto MyData nasce in Finlandia, all’università di Aalto, con alla base l’idea di rafforzare i diritti digitali dell’individuo. All’interno del contesto vi è anche la stringente legislazione europea sulla protezione dei dati personali, e la volontà di offrire alle aziende un nuovo modo di approcciarsi al cliente, basato non su incomprensibili “Termini e Condizioni” ma sulla fiducia reciproca.

Attualmente, grandi quantità di dati vengono continuamente raccolti, senza che se ne sfrutti pienamente la ricchezza dell’informazione. Il loro percorso e le operazioni di processamento a cui vengono sottoposti sono sempre più numerose, e aumenta la difficoltà nel tenerne traccia. Inoltre, l’utilizzo di software proprietari limita le possibilità di analisi dei dati a causa della scarsa interoperabilità fra le soluzioni adottate. Infine, va anche considerato l’impatto che questo modus operandi ha sugli utilizzatori finali: secondo un sondaggio riportato dal World Economic Forum, “Fully 78% of consumers think it is hard to trust companies when it comes to use of their personal data” (Orange, The Future of Digital Trust, 2014).

MyData cerca di trovare una risposta a questi temi attraverso un cambio di paradigma che ponga l’utente al centro, rendendolo capace di gestire i propri dati personali e di comprendere l’uso che ne viene fatto. Allo stesso tempo, la qualità dei suggerimenti ricevuti in base ai dati raccolti aumenterebbe, grazie alla maggior disponibilità degli stessi. L’elaborazione dei dati potrebbe portare anche a conoscenze più approfondite del proprio comportamento da utenti (*self tracking*), o alla ricezione di un compenso per il processamento dei propri dati (*data monetization*).

L’approccio a livello infrastrutturale permetterebbe l’indipendenza da specifici settori (sanità e salute, finanza, ecc.), favorendo una completa portabilità dei dati. Il rispetto e l’aderenza alle leggi verrebbe realizzato dall’infrastruttura stessa, permettendo alle aziende di sviluppare i software in maniera meno vincolata, acquisendo al contempo la fiducia del cliente grazie alla trasparenza e all’affidabilità garantite da MyData.

Nel complesso, il paradigma MyData si contrappone all’attuale standard, Big Data, proponendo nuove soluzioni che non perdano di vista la realtà di mercato in cui il software viene usato.

# 2 Smart Mobility

## 2.1 Mobility as a Service

MyData ha trovato un primo banco di prova all’interno del concetto di Mobility As A Service, anche grazie al coinvolgimento del Ministero dei Trasporti. Il progetto finlandese ha ricevuto un notevole slancio grazie alla tesi di Sonja Heikkilä, nella quale l’autrice suggerisce un nuovo concetto di mobilità per la città di Helsinki, in modo da rispondere alle sempre più frequenti sfide che vengono poste al settore dei trasporti. Le due principali difficoltà risiedono nell’inaffidabilità dei mezzi pubblici, spesso però l’unico modo per raggiungere determinate mete, comparata con la difficoltà di usare l’auto di proprietà, a causa della mancanza di parcheggi o degli ingorghi che si possono formare durante il viaggio.

Il paradigma MaaS descrive un nuovo utilizzo delle tecnologie applicato ai mezzi di trasporto, che propone il passaggio dall’auto di proprietà a mezzi di trasporto condivisi. Non si tratta solo di una migliore gestione dei mezzi pubblici, che proponga ad esempio soluzioni di pagamento online o potenziamento delle corse in base alla richiesta: Mobility As A Service si applica anche a mezzi come taxi, biciclette o sistemi di car sharing.

Questo cambiamento permetterebbe di aumentare l’efficienza con cui i mezzi di trasporto sono usati, eliminando gli sprechi che inevitabilmente derivano dal possesso di un autoveicolo e dal suo inutilizzo. L’accesso del privato ai mezzi di trasporto avverrebbe, in questa nuova ottica, attraverso un software in grado di calcolare una ottimizzazione per i mezzi condivisi, ad esempio raggruppando gli utenti per fasce orarie, tratte comuni e generiche preferenze. Grazie inoltre ad una gestione comune di mezzi di trasporto diversificati fra loro, la pianificazione del viaggio può comprendere tratte percorse in modalità diverse (ad esempio automobile e bicicletta).

Da questa proposta è nata MaaS Global, una startup (?) finlandese che, durante l’autunno 2016, rilascerà l’app per smartphone “Whim”, con la quale sarà possibile muoversi all’interno della città di Helsinki secondo il paradigma Mobility As A Service. Inizialmente, si potrà fare uso di mezzi di trasporto quali il trasporto pubblico della città, i taxi e le auto a noleggio, ma saranno presto integrati anche i servizi di bike e car sharing.

### 2.1.1 Mobility Profile e Journey Planner

Come *proof of concept*, sono state realizzate dall’università di Aalto due applicazioni, rilasciate su piattaforma Android e iOS, Mobility Profile e Journey Planner.

Mobility Profile raccoglie i dati personali dell’utente, li mantiene in un database relazionale e svolge le operazioni di calcolo del prossimo viaggio più probabile. Questa applicazione funziona come base di appoggio per Journey Planner, che riceve i suggerimenti calcolati e restituisce un feedback al processo sottostante, in modo da migliorarne l’accuratezza. Le API di comunicazione fra le due sono requestSuggestions(), requestTransportModePreferences(), e sendSearchedRoute(Place startLocation, Place destination).

In questo esempio, Mobility Profile non rispetta pienamente le specifiche dettate per MyData, ma è comunque possibile riconoscerne alcune caratteristiche all’interno del progetto. Ne sono un esempio la richiesta esplicita di un permesso (revocabile in ogni momento) per l’utilizzo di dati personali come gli impegni del calendario e lo storico delle posizioni GPS, e anche lo sviluppo di una applicazione separata per l’utilizzo dei risultati (Journey Planner) rispetto a quella che raccoglie i dati e li processa.

## 2.2 Smart Mobility for All

Rimanendo sul filone della mobilità intelligente, si sviluppa in modo indipendente da MaaS il progetto Smart Mobility for All.

L’idea centrale su cui si fonda la piattaforma SMAll è di costruire un insieme di servizi di mobilità, potenzialmente anche per mezzi di trasporto molto diversi fra loro, che gestisca l’intero ciclo di vita di un viaggio, dalla prenotazione all’effettivo spostamento. Si tratta di Smart Mobility poiché attraverso la raccolta e l’analisi dei dati di un utente è possibile fare inferenze sui prossimi spostamenti, proponendo anche l’acquisto dei titoli di viaggio più adatti alle abitudini e alle necessità dell’individuo.

Ancora in via di sviluppo, si basa sul concetto di modularità, che punta a fare di ogni servizio un modulo separato e indipendente dal resto. In questo modo si applica efficacemente il principio di suddivisione delle responsabilità, e migliora la manutenibilità sia del sistema nel complesso che dei singoli servizi.

All’interno dei vari componenti dell’infrastruttura si potrebbe prevedere un gestore di dati personali, del quale si porta in questa tesi un esempio semplificato. Anche in questo contesto risulta evidente l’importanza della gestione dei dati personali, non solo in termini di protezione della privacy ma anche con uno sguardo all’interoperabilità e al corretto funzionamento di un insieme di servizi eterogenei.

# 2 ½ GDPR??

# 3 Architettura MyData

Al fine di comprendere appieno le scelte effettuate all’interno del progetto, si mostrano brevemente le componenti e la struttura del modello MyData.

## 3.1 Entità fondamentali

L’architettura di MyData si costruisce su quattro componenti di base: l’utente finale, detto anche *Account Owner*, l’operatore MyData, o *Operator*, e due generiche entità che, da una parte, “producono” dati, e, dall’altra, li “consumano”. Queste sono dette rispettivamente *Source* e *Sink*. Mentre i ruoli di *Account Owner* e di *Operator* sono generalmente statici, quelli di *Source* e *Sink* sono fortemente variabili nel tempo e si possono applicare anche ad entità molto diverse fra loro, poiché definiti con un alto livello di astrazione. Convenzionalmente, si può identificare un servizio come “consumatore” di dati, mentre l’account dell’utente può essere un “produttore” di dati personali. In MyData, è possibile altresì che un servizio occupi entrambi i ruoli, o anche che l’*Operator* stesso rientri in questa classificazione quando si trova a compiere operazioni sui dati.

Il ruolo di *Operator* comprende operazioni di vario genere, fra i quali vi sono la gestione degli utenti, dei servizi e delle interazioni che avvengono fra le due parti. Si occupa anche di gestire l’Audit Log di tutte le operazioni che coinvolgono tali interazioni.

## 3.2 Service Registry, Service Linking

Viene indicata con *Service Registry* quella parte dell’*Operator* che contiene un database di tutti i servizi registrati presso quell’operatore. Esso contiene anche la funzionalità di *Service Discovery*, utilizzata dagli utenti per trovare nuovi servizi da utilizzare. In particolare, gli utenti possono venire a conoscenza di un nuovo servizio tramite un suggerimento, calcolato in base a corrispondenze fra caratteristiche dell’utente e del servizio, oppure tramite ricerca diretta.

Ogni nuovo servizio che vuole essere utilizzabile all’interno dell’architettura MyData deve quindi sottoporsi ad una procedura di registrazione, al termine della quale, in caso di successo, viene inserito all’interno del *Registry*.

Durante questa procedura, il servizio deve fornire almeno una descrizione del suo comportamento in formato machine-readable e human-readable: la prima permette a procedure automatiche una elaborazione corretta di suggerimenti, mentre la seconda è rivolta direttamente all’utente finale.

L’iscrizione di un utente presso un servizio avviene tramite un processo chiamato *Service Linking*, in cui l’Operatore MyData si occupa di realizzare una identificazione mutua delle parti. Tutti i token e le firme digitali scambiate durante il procedimento sono espresse in notazione JSON.

Al termine del *Service Linking* viene prodotto un *Service Link Record*, necessario per ogni futura interazione fra l’utente ed il servizio.

### 3.2.1 Oauth 2.0

//Da completare

http://tutorials.jenkov.com/oauth2/authorization-code-request-response.html

## 3.3 Autorizzazioni e Consent

Come specificato dal GDPR, ogni tipo di operazione svolta sui dati personali di un utente deve essere stata autorizzata dallo stesso tramite un permesso, che acquista in questo contesto una valenza legale.

La funzione di un permesso, o *Consent*,è particolarmente rilevante: definisce quali dati possono essere utilizzati e in che modalità, e identifica le entità *Source* e *Sink* fra le quali avviene lo scambio. Il processo di *Service Linking* deve essere stato completato con successo affinché sia possibile fare richiesta di autorizzazione, e ciò viene verificato tramite ispezione del *Service Link Record*.

Il ruolo dell’Operator in questa situazione è quello di recuperare le informazioni corrispondenti al servizio presso il *Service Registry*: queste vengono presentate all’utente, che decide se acconsentire al processamento di un determinato insieme di dati da parte di un servizio specificato. È possibile per l’utente chiedere una ridefinizione delle richieste del servizio, ma non al di sotto del limite previsto per un corretto svolgimento del servizio stesso.

In caso la procedura di autorizzazione termini con successo, viene prodotto un *Consent Record* che contiene tutte le specifiche negoziate fra le parti, insieme agli identificatori dell’utente e del servizio. Questo viene memorizzato all’interno dell’account, ma è possibile che il servizio o l’Operatore ne facciano richiesta successivamente.

Per dare la possibilità ad un utente di ritirare il permesso accordato, un *Consent* ha tre stati possibili: *Active*, *Disabled* e *Withdrawn*. Il primo è lo stato standard per il funzionamento, in cui l’accesso ai dati è consentito; si hanno poi gli stati “disabilitato” e “ritirato”, in cui l’accesso è impedito. Nel caso in cui il permesso sia ritirato è necessario provvedere all’emissione di una nuova autorizzazione, mentre in caso di stato disabilitato, è possibile attuare un cambiamento di stato, riportandolo al valore attivo.

Questo protocollo di autorizzazione all’utilizzo dei dati rispetta quindi quanto affermato dal GDPR, poiché il permesso viene dato volontariamente e in modo chiaro. Non è ambiguo, è informato, grazie alla specifica delle risorse necessarie, ed è possibile ritirarlo in ogni momento.

### 3.3.1 Kantara Consent & Information Sharing Work Group

//Da completare

<http://kantarainitiative.org/confluence/display/infosharing/Consent+Receipt+Specification>

### 3.3.2 UMA

https://kantarainitiative.org/confluence/display/uma/Home

## 3.4 Personal Data Storage

Nonostante non venga dato un peso rilevante a questa componente all’interno dei documenti di MyData, non è possibile prescindere dall’esistenza di un database che mantenga tutti i dati relativi ad un *Account Owner*. Non si tratta infatti solo di dati personali, ma di ogni dato utilizzato da un generico servizio o ad esempio inserito volontariamente dall’utente.

Non è specificato se il Personal Data Storage (spesso chiamato anche Personal Data Vault) faccia parte dell’ecosistema dell’operatore: ciò è possibile, ma non sono da escludere implementazioni alternative che prevedono il salvataggio delle informazioni presso il dispositivo dell’utente.

L’accesso ai dati contenuti all’interno del Personal Data Storage è possibile solo in presenza di un adeguato *Consent Record*, e le modalità di accesso non vengono regolamentate in maniera dettagliata. Si parla infatti genericamente di “Data API”, con le quali un *Sink* ottiene (in caso di richiesta legittima) un determinato *Resource Set* da un *Source*.

# 4 Analisi e Design

Il progetto del gestore di dati personali è partito da uno studio delle specifiche MyData, analizzando ogni entità all’interno di uno studio di fattibilità. Le soluzioni adottate sono generalmente di portata “enterprise”, e come tali inadatte ad un progetto di tesi. Per questo motivo, ho cercato di seguire durante il progetto le stesse linee guida e i principi alla base di MyData, preferendo quando possibile una implementazione più semplice e adatta al contesto.

Da quanto emerso nei paragrafi precedenti, la presenza dei componenti di alto livello elencati in seguito è necessaria per la realizzazione del gestore di dati personali. Questi sono:

* Un utente finale, che disponga di un account presso l’Operatore MyData e di un account presso il servizio di calcolo del prossimo viaggio più probabile;
* Il servizio che realizzi la logica di business, ad esempio un servizio che calcola il prossimo viaggio più probabile (*Most Likely Next Trip*);
* L’Operatore MyData;
* Un Service Registry, presso cui il servizio è registrato;
* Un gestore dei permessi per l’utilizzo del servizio e l’accesso ai dati dell’utente;
* Un Personal Data Vault per la memorizzazione dei dati.

## 4.1 Accounting E Servizi

Al fine di gestire gli utenti all’interno del progetto, sono stati previsti due tipi diversi di account:

* Account presso MyData. L’utente vi memorizza i suoi dati anagrafici, come ad esempio nome, cognome e data di nascita. Permette di associare ad ogni utente un Personal Data Vault, ed è punto di riferimento per tutti i singoli account creati presso i servizi.
* Account presso un servizio (qui Most Likely Next Trip). È unico per ogni utente, e mantiene lo storico di tutti i permessi che l’utente ha emesso.

Per quanto riguarda il servizio Most Likely Next Trip, si è scelto di utilizzare un componente già pronto, sviluppato da Nicola Ferroni nella tesi “Un sistema di previsione degli itinerari per applicazioni di smart mobility”. Nel corso di questa tesi, si porterà un esempio di come è possibile far interagire le due entità, regolando lo scambio di dati in modo quanto più possibile aderente al modello MyData.

## 4.2 Operatore MyData

La realizzazione di un completo MyData Operator comporta la costruzione di diversi componenti. Le funzioni dell’Operatore comprendono la gestione del *Service Registry*, degli account utente e del processo di *Service Linking*. L’Operatore deve inoltre occuparsi della mutua identificazione fra le parti nel processo di autorizzazione del servizio, durante la firma dei *Consent*, e autorizzare il flusso di dati quando ne viene fatta richiesta.

In questo progetto si è preferito suddividere le responsabilità corrispondenti a queste operazioni in una molteplicità di classi, invece che un’unica entità, in conformità a principi di semplicità come il rasoio di Occam o l’acronimo KISS (*Keep It Simple, Stupid*). Pertanto, si realizzeranno separatamente:

* Un gestore di *Consent*, che si occupi dell’emissione dei permessi e del cambio di stato degli stessi in caso di richiesta da parte dell’utente.
* Un *Service Registry* semplificato, dove ogni servizio si registra per poter essere accessibile all’interno di MyData.
* Un’entità che metta in comunicazione il Personal Data Storage con il servizio, recuperando i dati necessari (punto di *policy enforcement*).
* Un gestore per le operazioni di autenticazione fra utente e servizio.

Per questo motivo, si sceglie nel seguito del progetto di non citare più il concetto di Operator, facendo invece riferimento alle entità particolari.

## 4.3 Service Registry

Una implementazione proposta dal team di sviluppo finlandese può essere trovata all’indirizzo ( <https://github.com/digitalhealthrevolution/serviceregistry> ) [ to fix ]. Tuttavia, la realizzazione di una applicazione sul modello di quella proposta esula dagli obiettivi di questa tesi a causa della sua complessità, pertanto se ne propone di seguito una versione semplificata.