MyData Architecture Framework

Release 2.0

Authors:

Harri Honko	Tampere University of Technology
Yki Kortesniemi	Aalto University
Jens Kremer	Aalto University
Samuli Tuoriniemi	University of Oulu

Table of contents

Notice

MyData Architecture defines the operations and APIs between the Operational Roles (Operator, Service, Source, Sink etc.). Any descriptions or figures of the role's internal structure or operations are for illustrative purposes only.

Acknowledgements

We want to acknowledge the contributions of the following people:

Anette Alén-Savikko	Aalto University / Helsinki Institute for Information Technology HIIT
Nomi Byström	Aalto University / Helsinki Institute for Information Technology HIIT
Harri Hirvonsalo	University of Oulu
Antti Kallonen	Tampere University of Technology
Kai Kuikkaniemi	Aalto University / Helsinki Institute for Information Technology HIIT
Tuomas Paaso	VTT Oy
Olli Pitkänen	Aalto University / Helsinki Institute for Information Technology HIIT
Antti Poikola	Aalto University / Helsinki Institute for Information Technology HIIT
Sari Vainikainen	VTT Oy
Jani Yli-Kantola	University of Oulu

1. Introduction

Applications and services collect increasing amounts of personal data about their users, and leverage it to extract valuable knowledge about them. This information can be used for providing new services and for profiling individuals, and the results are monetizable input for e.g. targeted advertising. Unfortunately, individuals themselves typically have little or no control over how their data is created or used.

This document presents a reference architecture framework for technical realisation of MyData principles, a human centric approach to liberate the potential of personal data and to facilitate its controlled flow from multiple data sources to applications and services. The simple core idea, the *individual in control of their own data*, is both a movement for digital human rights and an initiative for opening new business opportunities. It responds on a practical and technical level to individuals' growing demand for control over their personal data and to organizations need to fulfill the requirements of tightening data protection regulation. Though both of these goals can be achieved within the current legal framework, the lack of interoperable implementations has kept them mostly a distant goal - a situation that we now want to change.

The architecture aims to provide a standard for implementations that

- satisfy the legal requirements for processing of personal data and, thus, prevents unwanted and improper processing of the individual's personal data
- enable the individuals to easily grant and withdraw their consent for data processing
- enable the individuals to view and enter into contracts related to their personal data processing
- provide the individuals notifications about and means to object data processing not requiring their consent
- provide transparency to individuals about how their data is being used

In the following sections we present an overview of the GDPR and the identified use cases and roles with MyData architecture framework (Section 2), then an overview (Section 3) and details of the architecture: MyData Account (Section 4), Service Descriptions and Registration (Section 5), Service Linking (Section 6), Contracts (Section 7), Consenting (Section 8), Notification (Section 9), and Objection (Section 10). Each detail section has a corresponding detailed technical document (see References).

This document is part of the MyData Architecture Framework release 2.0. The reader is assumed to be familiar with the parallel technical specification documents available at https://github.com/mydata-sdk/mydata-docs/tree/master/architecture_specs .

1.1 Notational Conventions

The keywords "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

1.2 Terminology

Key terminology used in this specification is defined in the Glossary of MyData Architecture Framework release 2.0 available at https://github.com/mydata-sdk/mydata-docs/tree/master/architecture_specs .

1.3 Formats

In MyData Architecture, all data records and their respective digital signatures exchanged between actors are expressed using Javascript Object Notation (JSON). Digital signatures are expressed as JSON Web Signature (JWS) structures and cryptographic keys as JSON Web Key (JWK) structures.

In this document, JSON definitions of the data records are presented without JWS structures.
All Timestamps are in UTC in the NumericDate format as defined in [RFC7519].

2. Background: MyData, GDPR and Use Cases

This section first summarises the core concepts from MyData Whitepaper [1], then provides an overview of EU's data protection regulation GDPR, introduces the use cases for the architecture, and, finally, discusses key related standards. The goal is to help understand systems built according to MyData principles without going too deeply in the technical implementation details; technical documentation and code release of a reference implementation of a MyData architecture is provided separately [2-4].

2.1 Introduction to MyData

At the heart of MyData are Mydata Account and five operational roles.

The human centric concept in MyData architecture builds strongly on the ***MyData Account,*** which the individual uses to manage their personal data. It contains individual's digital identity or identities, linked services, and data use related policies and consents. Potentially, these are complemented with individual's other data that help in providing additional or improved services. The Account is hosted by an independent MyData Operator, which also provides the tools for managing the account.

Figure 2.1: Five operational roles within the MyData architecture include 1) individual as the Account Owner 2) MyData Operator, 3) Service, 4) Source, and 5) Sink. '6th' Dual-role Service is combination of the last two. The MyData Account is hosted by the Operator.

There are five key operational roles in the My Data architecture: *Account Owner*, *MyData Operator*, *Service*, *Source* and *Sink* as shown in Figure 2.1. Actors (organisations and individuals) may work in one or more of the operational roles. It is e.g. very typical for an organisation to simultaneously be in the role of Source and Sink.

These role names can be mapped to established authorisation and policy frameworks, e.g. with use of OAuth: Account Owner maps to resource owner (RO), Source service equals resource server (RS), Sink service equals client, and finally, MyData Operator is an entity that usually also provides the resource owner the needed identity management - such as OpenID Connect provider (OP) - and Authorisation Service (AS) using a combination of appropriate OAuth 2.0 profiles.

Account Owner is the individual who created and is using the account to e.g. link new services (see *Service Linking*) and consent to data use either within a service itself or re-use via 3rd parties (see *Consenting*), agree to a contract about data processing, object to the processing, and receive notifications about realised data processing at services. Account owners are always the individuals whose personal information is processed and who manage the use and permissions for personal data processing. One account owner may have multiple accounts.

Service is the party processing the personal data of an Account Owner in order to provide specific services. These services can be provided in context of the Account Owner's personal service, someone else (e.g. a request to view or use Account Owner's personal data held by the Service), or secondary use by the Service's linked and trusted third parties (e.g. as part of mass data transfer by a linked public government registry or research study). A Service that only processes an Account Owner's personal data for its own purposes may employ the MyData infrastructure to inform or communicate with an Account Owner.

In addition to processing personal data, Service may also implement the roles of Source and/or Sink for the same personal data or its subsets it processes:

Source is a Service entity that can, when authorised, provide access to an Account Owner's personal data for one or more Sinks.

A ***Sink*** is a Service entity that can, when authorised, fetch data from one or more Sources and use (process) the data for the agreed purposes.

Both Sources and Sinks need to provide the corresponding MyData-compatible interfaces. Source interfaces enable the management of data provisioning, while Sink interfaces enable the management of data usage. It is quite common that a service is working both as Source and Sink, and therefore provides both Sink and Source capabilities and interfaces.

The ***MyData Operator*** is the entity that provides, hosts and manages MyData accounts and their user interface. An Operator also provides the underlying mechanisms for linking Services, Sources and Sinks

to the account, and managing the account specific policies, notifications and authorisations, ie. it works as a privacy dashboard and intermediary for the individual. The basic vision of the architecture enables the existence and use of multiple operators. Each individual can choose to use one or more operators to manage their privacy through specific policy authorisations. In addition to processing personal data, operator can also implement the roles of Source and/or Sink for the same personal data or its subsets it processes as controller. It can e.g. provide the Account Owner's public privacy preference profile to be shared through an internal authorisation function.

2.2 GDPR, the Legal Framework for Personal Data Processing in EU

The legal framework for protecting personal data in Europe is based on strong fundamental rights protection. That is because the unlimited collection and use of information about individuals can have negative consequences for individuals and for the whole society. Furthermore, the right to privacy and the right to data protection have developed on a global scale in order to address the challenges deriving from digitisation and technology.

Consequently, in EU all personal data collection and processing interferes with individual rights. Personal data processing therefore needs to be adequately justified. Firstly, this means that all data processing operations have to be lawful. They need to be based on a specific ground provided by a relevant law and they can't be 'illegal', and therefore against a relevant law. Secondly, all personal data processing operations are required to adhere to the general principles of data protection (for example the principles of limited collection or purpose limitation) as outlined in national and international data protection conventions and regulations.

For MyData architecture framework documentation, we use the EU General Data Protection Regulation (GDPR). The GDPR is one of the most modern privacy laws that regulates privacy standards throughout the many countries in Europe. Furthermore, the EU privacy regulation is based on the fundamental and human rights to privacy and the protection of personal data, and relies therewith on globally valid rights and principles. Additionally, the GDPR applies beyond the borders of the EU if EU citizens are offered goods or services or if persons in the EU are monitored. As a consequence, GDPR compliance is a global topic and the set standards are relevant way beyond the borders of the EU.

The GDPR defines a number of legal roles relating to personal data processing, most importantly the data subject, the controller, and the processor. Each role is subject to differing rights and obligations. The notion of personal data is defined as "any information relating to an identified or identifiable natural person" (the data subject) (Art. 4 (1)). Data controllers and data processors are either natural persons or legal persons, public authorities, agencies, or other bodies (Art. 4 (7)-(8)). A data controller is the entity which "alone or jointly with others determines the purposes and means of the processing of personal data" whereas the data processor "processes personal data on behalf of the controller". The notion of processing is wide: according to Art. 4 (2) GDPR, processing signifies "any operation or set of operations which is performed upon personal data or sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction".

In order to be lawful, every processing of personal data requires a legal basis. This means that all data processing activities are required to be based on a clearly defined grounds for processing. The GDPR presents an exhaustive list of such possible grounds: Data processing can only be legal if it is based on an individual's consent, necessary for the performance of a contract, justified in the legitimate interest of a controller, as well as if processing is in the vital interest of the data subject, public interest, or for the fulfillment of a legal obligation. From the perspective of informational self-determination of the data subject, consent signifies an especially important legal basis. According to the GDPR, consent means an indication of the data subject's wishes by which they signify agreement to the processing of their personal data, either by statement or by "clear affirmative action"; consent must be freely given, specific, informed and unambiguous.

Article 7 of GDPR provides the framework for consent: Firstly, the controller must be able to demonstrate the existence of consent. Secondly, in the context of written declarations containing also other matters, consenting must be clearly distinguishable, accessible and understandable in order to be valid. Thirdly, the data subject can always withdraw their consent and this must be as easy as consenting. Fourthly, in assessing the free nature of consent, particular account is to be taken of whether the performance of a contract, including the provision of a service, is made conditional on the consent to processing of unnecessary data (i.e. that the processing is not necessary for the performance of the contract). As consent as a legal basis is embedded in the overall regulations in the GDPR, it's practical use must cohere with the overall rules and principles set forth in data protection regulation, such as for example transparency, purpose limitation or limited collection. The GDPR also requires parental oversight for children's consent (below the age of 13-16, depending on national definition) in cases where information society services are directly offered to children. Furthermore, some special categories of personal data require stronger justifications, including explicit consent for processing.

It is important to emphasize that the rules for valid consent have significantly tightened with the GDPR. Consent should not be used when individuals have no genuine free choices, for example where there are significant power imbalances between a data subject and a controller, or where consent to personal data processing is made conditional for the provision of a service. It is furthermore important to understand that consent in the GDPR is separated from agreeing to a contract. If a contract is made, personal data can be processed only in so far as the processing is necessary for the performance of the contract.
Another option for processing personal data on a lawful bases is by referring to a legitimate interest of a data controller. This means that personal data can be processed when there is a present, lawful and clearly specified interest of a controller and the personal data processing does not unproportionally interfere with individuals' rights. Legitimate interest requires sound justification and reasoning as well as a sophisticated balancing test on the side of the controller.

Personal data can also be processed if this is of vital interest (in life-or-death scenarios) to the data subject or other natural persons. Furthermore, personal data can be processed by a public authority in case of a justified public interest.

The MyData infrastructure takes into account the different legal bases and requirements for personal data

processing stemming from the GDPR. It furthermore fosters compliance with data protection laws and regulations by enabling individuals to exercise control over their personal information. Additionally, the MyData infrastructure assists data controllers with privacy compliance and supports communication between data controllers and data subjects. In that sense, the MyData infrastructure acts as an intermediary and as a tool of communication between data subjects and controllers. It remains, of course, the task of the data controller to fully adhere to its obligations deriving from data processing rules and regulations.

2.3 Use Cases

In this section we present use cases to illustrate the use of this architecture framework for personal data management under different (legal) situations covered by GDPR. The section contains six use cases over contract-based processing, explicit consenting and processing related notifying or objections. A consent can cover processing of data by another service where the same person has a relationship with (‘my data for my needs in my other service’), by another person or by a third party organisation (in any case, a legal entity). Processing notifications can be provided for all use cases.

Note that use cases will be more straightforward in presence of a single-sign-on service between involved parties, but cases are doable also with service specific credentials.

Contract

- Account Owner enters into a contract with a Service. The Service processes personal data that is necessary for the fulfilment of the contract. Account Owner’s account at the Operator may be used for documenting the presence of a contract between the two parties if this is supported by both Service and Operator.

Consent-centric

- **Consenting:** Account Owner gets a proposal from an earlier contracted Service or signs up for new Service and issues a consent to process personal data for one or more of potentially many specific purposes defined and offered by the Service.
    - **Consent adjustment:** A Service is processing personal data for a specified purpose or set of purposes for which it has received a consent from the Account Owner - at some point the Service suggests processing of the existing personal data for a new and significantly different purpose, or it introduces new earlier unknown third parties onto its processing chain.
    - **Consent withdrawal:** The Account Owner withdraws from an earlier issued consent to process personal data.
- **3rd party re-use:** With Account Owner’s consent, a Sink accesses personal data from the Source and processes it for the defined purposes. This consent can also be adjusted or withdrawn.

Other

- **Processing notification:** A Service processes personal data using legal ground other than individual's consent and notifies the Account Owner about the processing to meet its public or own requirements for data processing transparency.
- **Objecting to process:** Account Owner's personal data is being processed and the Account Owner wants to object to this processing.

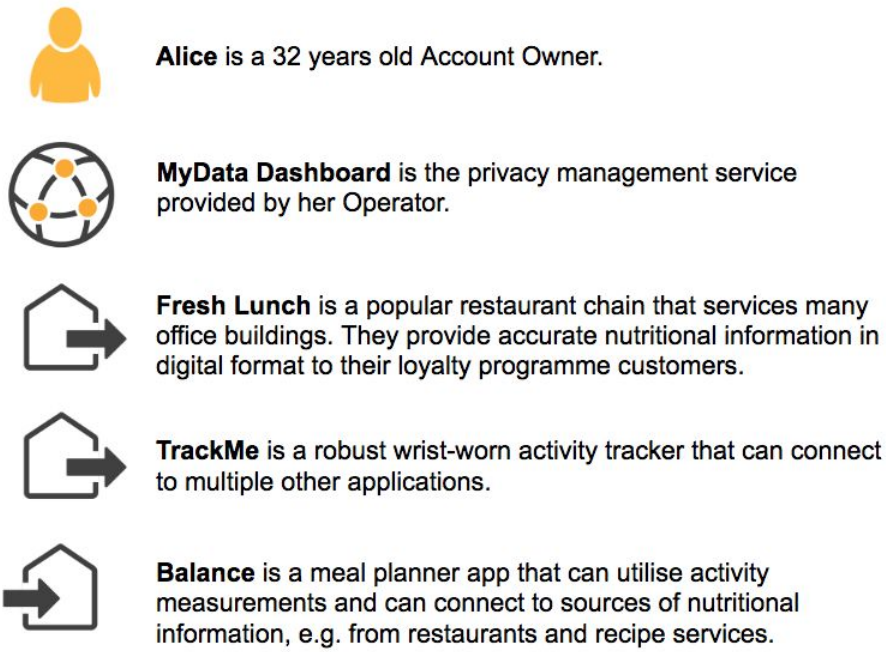


Figure 2.2: The parties in the use cases.

The parties of these use cases have been presented in Figure 2.2 and the prerequisites for the use cases are as follows:

- Fresh Lunch, TrackMe and Balance are already known services at MyData Dashboard
- Alice has an account at MyData Dashboard
- After signing up to TrackMe through the contract use case Alice has the TrackMe service listed on her account at MyData Dashboard

2.3.1 Contract Use Case

A *contract* use case documents the personal data processing agreement between Alice and TrackMe to her account at MyData Dashboard.

Story for contract:

When Alice signs up to paid subscription based TrackMe she agrees to the basic data processing requirements of the service. The contract allows only a minimum set of personal data processing for fulfillment of basic service features such as her address and credit card details. Additionally TrackMe may well claim as part of the contract the details on personal data processing that is based on their legitimate interest. TrackMe holds the original contract but they are providing contract copies to customers with MyData Accounts as a convenience service.

TrackMe has made an agreement with MyData Dashboard to provide data processing transparency to its customers. In light of that, Alice can see in her MyData Dashboard account that TrackMeCorp is processing her credit card information and home address in order to receive payments and send billing receipts by mail because such personal information is necessary for fulfilling the contract to provide the service.

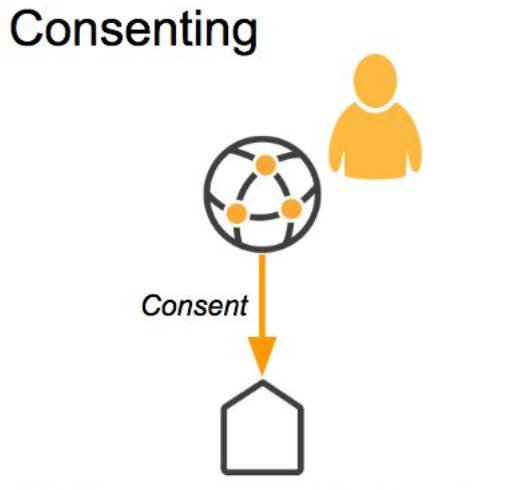
2.3.2 Consent-Centric Use Cases

The *consenting* case is a simple scenario where *Alice* (the Account Owner) authorises *TrackMe* to process her data within the service's own defined purposes through providing her consent.

The *consent adjustment* case covers a scenario where the existing service TrackMe introduces a new significantly different internal or third party processing related feature that necessitates Alice to provide or to adjust her earlier consent to cover a new data processing operation.

The *consent withdrawal* covers a scenario where Alice withdraws her earlier issued consent via the MyData Dashboard. There's no detailed use case walk-through for this.

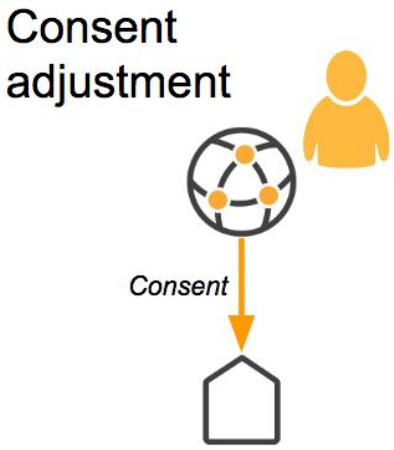
The *3rd party re-use* case covers a scenario, where Alice authorises a new service (in our exemplar *Balance*) to access her data from other compatible services (*Fresh Lunch* and TrackMe) using her Operator (*MyData Dashboard*).



Individual issues a *consent* to the **Service** authorising it to process personal data for a specific purpose.

Story for consenting:
Alice has already a contract with TrackMe but that doesn't cover the detailed personal data TrackMe would need for providing their newly introduced and marketed advanced analytics to Alice. When entering the advanced features section in TrackMe's 'What's new' walk-through Alice finds the new HR tracking features so interesting she decides to activate those. At this stage TrackMe asks her consent for processing the required personal data from compatible wearables and explains clearly for which end result the data is used for. Alice has earlier *linked* TrackMe with her account at the MyData Dashboard and thus the resulting consent is recorded there also.

Story for consent adjustment:
Alice has been an active triathlon trainer for last 18 months, and she's been waiting for her favourite tracker to release their new strapless HR monitoring feature with the new hardware version of the wrist tracker - and now the product is finally out just for the spring season sales.



Individual issues a new *consent* to the **Service** authorising it to process personal data for a new purpose the previous consent didn't cover.

The new device and its companion app come with advanced TrackMe's internal tools for evaluating a person's cardiac condition and e.g. elevated risk of over-training that could cause permanent damage to trainer's cardiovascular system if not reacted to in a timely fashion. To enable this feature TrackMe has had to introduce a new licensed technology to their data processing pipeline. Earlier related analytics was beyond the TrackMe's service capabilities, though it has been able to link with authorised simpler HR data summaries from external tracking apps.

Alice has earlier approved the basic HR data processing terms and purposes of TrackMe (a consent record hosted at the MyData Dashboard works as the proof of that approval, *consenting*). Now due to the significance of the new feature in TrackMe product package the service triggers a *consent adjustment request* for its existing customers (for new users, the new functionality related purpose can be included in the consent from the beginning).

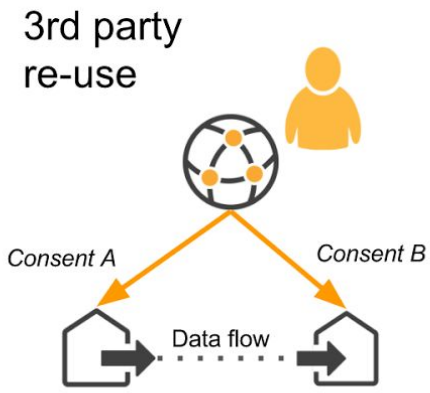
With traditional means, reaching out to the existing clients such as Alice to ask their consent for extended data processing is the burden to the TrackMe company, the data controller. As Operator, MyData Dashboard offers its registered services a practical terms and consent adjustment service (including push notifications to TrackMe's MyData Dashboard -linked customers), so this process can now be taken care of with ease.

As TrackMe's new data processing purposes have been updated, the specifics have been explained in a few human readable text sentences for the web tool, and the consent adjustment option has been activated at MyData Link by the TrackMe's admin user. Now, Alice gets a notification explaining that her full use of new tracker device requires an action at MyData Dashboard.

Alice enters the operator's page where the TrackMe request welcomes her to the landing page. After tapping on the open request she gets a full view of new extended data needs of the service, and gives her explicit consent which is again stored by the parties for documentation and proof purposes. Off she goes for the spring's first swim in still cold waters of lake Erie to get a nice peak on her HR data that gets noted

down on TrackMe's measurement database and her cardiovascular data gets analysed by their new analytics tool.

Story for 3rd party re-use:
Alice gets a tip about a new application "Balance" for her mobile phone from her personal trainer Bob and decides to try it out.



Individual issues a pair of consents: *Consent A* authorises the **Resource Server** to provision personal data to the Client, and *Consent B* authorises the **Client** to process that data.

Balance is MyData compatible and, thus, it asks Alice, if she already has a MyData Account with which to link, or if she would be interested in creating a new account with a recommended operator. Alice picks the 'Link existing account' choice and authenticates to her account at MyData Dashboard from within the Balance app.

The Balance app then suggests that Alice links also her activity tracker and restaurant bill data, as access to these would improve the use of the app. However, Balance can be used also without these enriching data sources. By using the embedded MyData Dashboard user interface within the Balance app, Alice *authorises* the Balance app to access her data from TrackMe, but leaves the restaurant bills out.

Later Alice logs in to her operator MyData Dashboard and sees that the recently connected Balance app now appears on the list of her services. User Interface of the MyData Dashboard has functionality for *discovering* compatible services amongst those the operator has on its service listing. With that Alice finds out that Balance app could also use data from the Fresh Lunch restaurant chain. Alice decides to subscribe to Fresh Lunch's loyalty program, link Fresh Lunch to her account at MyData Dashboard, and authorises the Balance app to access her restaurant bill data.

After authorizations are granted, the Balance app fetches the data from the API interfaces offered by TrackMe and and Fresh Lunch using a *data connection*. Balance app then uses Alice's data to provide infographics of her nutritional behaviour versus her current health state and targets.

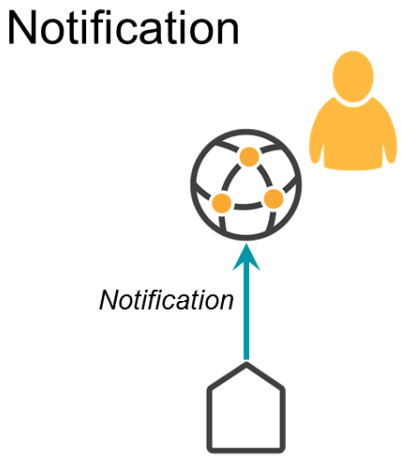
2.3.3 Other Use Cases

Notification case illustrates a mechanism to keep Alice well-informed about processing that happens without an explicit consent (e.g. due to local legislation). Notifications can be understood as a subscription service to more or less frequent information flow about Alice's data use.

Objection use case builds on the same situation where processing of Alice's data happens without an explicit consent, but in this case she is offered the option to actively object to the processing.

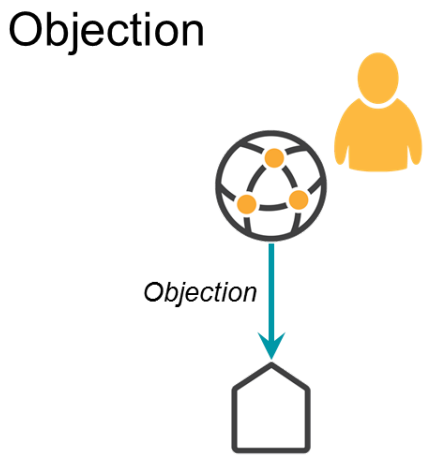
Story for notification:
Alice's EHR data has been utilised several times over the past months by various clinical organisations as her recent knee surgery resulted in multiple consultations with various physiotherapy and general practitioners and X-ray and magnetic resonance (MR) images of her foot needed to be circulated among the healthcare professionals.

MyData Dashboard has made agreements with the government and various counties in its operational region to provide data access management and processing reporting for public services. One such service provider is the local district hospital that carries out secondary care for region's residents.



Service *notifies* the **Individual** that personal data is being processed.

During and after the the knee rehabilitation process Alice receives to her MyData account a processing notification from MyData Dashboard stating clearly which entities and who particularly within the organisations have been processing her data. This notification contains, for example, that the hospital system has stored information on her surgery, including X-rays and the type and amount of anesthetics used during her surgery and which doctors have accessed this information at what time.



Individual issues an *objection* refusing the **Service** to process personal data.

Story for objection:
The same EHR data and genome data (derived from a separate bio-banked blood sample) of Alice is used for national healthcare outcomes research and for separately screened clinical research by researchers in public and private organisations. Actually her country's specific healthcare regulation sets this as an implied legal right of the organisations (public interest as the legal basis).

Ongoing research use on her data is informed as in notification use case above to her MyData Dashboard account, and she tends to check the purposes and targets of these research projects - most often she's positive with her data being used for common good of the mankind.

One particular new project by a pharmaceutical company has declared a purpose for their research that although approved by the research council's assessment Alice doesn't agree to - due to her ethical opinions we won't open up here - and she wants to withdraw from being a data donor for such research. MyData Dashboard provides an 'I don't want my data to be processed' control on the Alice's data control view next to the informed data processing listing, and thus Alice can provide an *objection*. Once activated, Alice's related data resources are excluded from future processing.

3. MyData Architecture Framework

This section gives an overview of the key functions in MyData Architecture Framework. For each function, one or more alternative implementation can be created, and those are referred to where and when an exact function mapping exists.

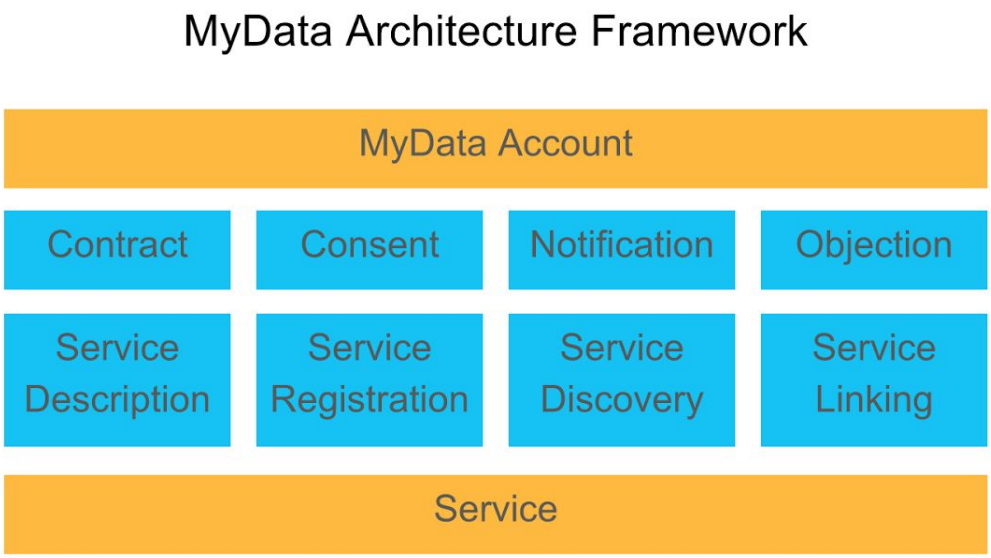


Figure 3.1: The elements of MyData Architecture Framework

The elements of the Framework have been shown in Figure 3.1. For the individual, the Framework appears mostly as the *MyData Account* which both stores all the relevant information about and offers the tools for effective data management.

For the *Service*, the Framework involves implementing a number of APIs to be able to communicate with the MyData Account and other Services, if the service functions as a Source and/or Sink. Service must also describe it's MyData capabilities by creating a *Service Description*. The Operator can also assist in discovering interesting services both for service developers and individual by supporting *Service Discovery*.

To start to manage the data usage of a particular service, the individual first has to *link* the service to their MyData Account. Only after that, are the *Contract*, *Consent*, *Notification* and *Objection* functionalities available.

All of the elements have been described in more detail in the following sections, which also refer to the existing reference implementations of each element.

It should be noted that certain parts of framework-compliant solutions can be implemented with alternative technology selections - such as access authorisation for the 3rd party re-use can well be implemented with external OAuth or UMA (or any other legacy or novel technology) based authorisation architecture. It will just be complementing the layers needed to comply with this framework.

4. MyData Account

Functionally, MyData Account is the key enabler for the individual in being informed about and managing the processing of one's personal data: getting notified about or objecting to data processing as well as authorising, controlling and logging the data processing within a service or data flows between multiple services. It keeps track of individual's consents, notifications, objections and other relevant information in a single place thus providing a unified view to all the personal data processing transactions and authorisations.

Typically, MyData Account contains

- basic information about the Account Owner, e.g. contact detail, preferences etc
- individual's identities for both single sign-on (SSO) uses and for managing the service links and policy documents. Depending on the implementation, account could hold just the public keys or both the public and private keys. Naturally, hosting private keys poses more stringent security requirements for the Operator.
- individual's linked services and their policy documents.

To be operational, the account has to be hosted by a MyData Operator. It is expected that most of the Operator services will be provided by organisations, though it is also possible for individuals to run the operator software themselves thus becoming self-operators. The key difference is that organisations can potentially provide additional trust, assurance and security levels compared to a self-operator. Acceptance and audit process for an organization wishing to serve as a trusted operator are beyond the scope of this document[1].

The information in a MyData Account is designed for portability allowing the individual to switch operators by taking their account with them. The implementation of portability as a service between Operators is deferred to a later architecture release.

All transactions performed within the architecture framework are recommended to be recorded into an audit log. This log can be can be used for auditing purposes (c.f. the current databases used in hospitals that automatically log each time someone accesses medical records so that any unauthorised use can later be acted upon) as well as constructing an up-to-date summary of all service links, consents, data transfers, notifications, objections and contracts associated with a specific MyData Account. Exact audit log implementation is left for implementors of Operator and Service instances as implementations will reside on multiple different platforms and OSes.

Example of the MyData Account as implemented in the MyData SDK can be found in the Account Service specification [2].

---

[1] It is likely to be defined along extending the criteria set for entering e.g. a regional or governmental trust network - ISO/IEC 29115 entity authentication assurance level or FICAM TFS are examples of trust and assurance level related requirements that may apply.

5. Service Descriptions, Registration, and Discovery

Each service compatible with the framework MUST maintain an up-to-date *Service Description* detailing the purposes and capabilities of a service, which reflect its role and applicable use cases. Some services are provisioning out personal data, some handle it only internally, some are willing to connect to existing external data providers to use data, and so on.

Service descriptions form the bases of *service registration*, which provides two major functions: it maintains a 'virtual' database of all services and operators (knowledge of configuration parameters and endpoints of the services and operator's respective information) and it enables searching for compatible services and operators (*service discovery*), both for the Account Owners using the services and for the developers of other services. Registration also provides each service a unique ID within that (virtual decentralised or static operator-centric) registry, which the service MUST have for many operations such as Consenting.

This section first introduces the different service descriptions and then goes over the Service Registration and (briefly) Service Discovery processes. A more detailed description of the service descriptions is available in the Service Descriptions Specification available at https://github.com/mydata-sdk/mydata-docs/tree/master/architecture_specs .

5.1 Service Descriptions

For efficient service management, discovery, and matching, each service MUST describe itself in the form of structured Service Description, which is published to the world through its /.well-known/MyData/configuration endpoints. Prerequisite for global service discovery is that MyData services need to announce their well-known locations appropriately per RFC 5785.

In addition to the compulsory description data such as service ID and some promotional material e.g. service logo, a Service Description is an aggregation of multiple levels of descriptions (see also Table 5.1):

- *MyData Configuration* describing the URLs to be used in defined MyData transactions
- *Human Readable Description* forms a basis for promoting and introducing the services to Account Owners
- For Source services, *Resource Description* details both the data the service is able to provision and the service APIs for accessing the data. Mapping to e.g. OAuth practice this would be equivalent to registering the resource server specific scopes.
- *Purpose Descriptions* are required for building detailed processing notifications, consent proposals for Account Owners, and for intelligent service discovery and matching

Table 5.1 Matrix of applicable descriptors per Framework's use cases. The phase where a particular descriptor is needed, and the relevant MyData configuration endpoint are provided below the 'X'.

	Consenting	Re-use of data	Notification	Objection
MyData Configuration	X Service registration	X Service registration	X Service registration	X Service registration
Human Readable Description	X Linking	X Linking	X	X
Resource description	-	X Authorisation	-	-
Purpose description	X Authorisation	X Authorisation	-	-
Processing description	-	-	-	X
Notification description	.	.	X	.

5.1.1 MyData Configuration Description

MyData Configuration description lists the particular description URIs, which may point to data hosted by the Service or some other entity in the Internet.

5.1.2 Human Readable Description

Human Readable Description consists of textual presentations of the service, and possible material (logo etc.) for promoting the service. These are used in various stages for presenting the service and its data in a non-technical, end-user friendly way, such as in the service store, where the Account Owner can discover services.

5.1.3 Resource Description

Each source service must provide a description of the data it is processing or capable of provisioning to other (sink) services. This description presents the maximum possible data set provided by a Source, though individual Account Owners may have or be willing to share only a subset of the data types. Service developers use the resource description for finding and integrating relevant Sources to their Sink. Examples are provided in a separate Service Descriptions Specification.

Source services must also provide the technical information of the API and endpoints to be used for accessing a particular resource. This description can be, for example, a JSON-LD/Hydra or YML/Swagger document presenting the API of a REST-style resource interface. Example description for our use case's TrackMe's heart rate resource is provided in Swagger 2.0 (API) format at [ref TBA].

5.1.4 Purpose Description

Purpose Description presents for what purposes and under what processing bases the Service processes the data, and which data items are used by each purpose. This information is used for informing of data processing under different legal bases and especially in constructing the consent proposal in consent-based use cases. Furthermore, this data may be used for finding compatible services.

Linking of data and purpose descriptions to Linked Data and common classification schemes is strongly recommended. The purpose is to establish a common understanding of various data elements, to enable correct interpretation of the data, to enable searching of data based on different criteria and relations of data elements, and to enable utilisation of data in new services and applications.

5.1.6 Notification Description

Notification description details, what types of notifications about processing the service is capable of delivering.

5.2 Service Registration

Service Registration is a three-step process, where the necessary information for using and discovering the service are published in the registry in a uniform way as shown in Figure 4.2:

1. Service provider registers the services and receives a unique ID (within this registry) for the service. Many operations such as Consenting or data access logging require the services to be identified.
2. Service provider adds the service descriptions discussed above
3. Service Developer provided URLs of service endpoints are linked with the service Resource Description.

The service registration may be offered as a service that links the registrations to Operators' service catalogues. Some operator may choose to provide a private service registry if their use cases don't demand cross-operator service visibility.

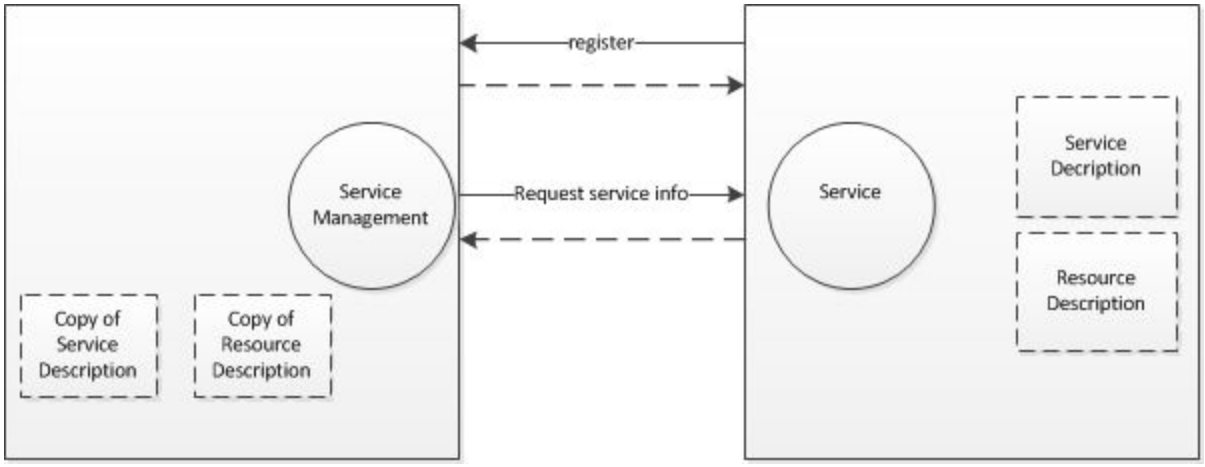


Figure 4.2: Registering a service to a Service Registry

Service Discovery

Service Discovery can be used for many purposes - e.g. to find and link relevant services to one's MyData Account, to recommend relevant services to an Account Owner, and to find relevant Sources for various personal data driven application purposes. As an example, a Sink (application) needs to find Sources that provide certain types of data, e.g. grocery data or fitness data that are needed as input for its internal analysis or processing.

Basic discovery is enabled by the Human Readable Service Description and additional information (such as service provider -documented lists of compatible services) associated to the description. Resource Description supports Source service discovery e.g. based on information readable from description endpoint (see https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs)for detailed description format) and linked API documentation. Service Description's *Processing Descriptions* document the applicable legal bases for a service. This data is available from the */.well-known/mydata/servicedescription* endpoint.

For more intelligent service discovery and, in particular, intelligent service matching, a semantic service data description is preferred. Semantically enriched descriptions bring support for multilingual searches, for matching different data elements describing the same thing, and for using relations of data elements in searching.

Semantics also support interoperability between different services. It enables more automatic mapping of Source data model to Sink data model. As within the services used in the current Internet there rarely are any semantic descriptions provided, which means for now more manual integration and testing work for any service's developer. For example there may be several Sources for fitness data and every one of them has their own format to describe the fitness data and different set of properties available.

This version of Framework does not provide a more detailed specification for service discovery.

6. Service Linking

Service Linking is the action, where a service is linked to the Account Owner's MyData Account. For the service to agree to the linking, the Account Owner has to be able to sign in to their account at the service or otherwise prove that they control that account and are, therefore, allowed to create the link.

With services that the individual uses, they can also initiate the linking process. However, some services (e.g. governmental ones) process personal data based on a legal bases other than the individual's consent, so the individual might not even know such processing takes place. Yet the services could provide notifications of the processing to the individual, or even allow objecting to the processing if the service is linked to the individual's MyData Account. In such cases it would be convenient, if the individual could log into e.g. a governmental portal (say, suomi.fi) and request that all registered services send their notifications to the individual's MyData Account, which would then initiate automatic linking of all services to the indicated MyData Account. Nevertheless, automatic linking leaves it up to the Account Owner how detailed information they wish to receive from that service. Also, as with any other service, the Account Owner is free to move the Service Link to another Operator.

Only after a Service Link has been established, the Account Owner is able to manage the data sharing and processing authorisations, notifications and objections for this service. Service Linking process is documented in detail in Service Linking Specification available at https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs.

Any service to be linked needs to be discoverable and available for dynamic registration per the processes of Section 5 above. Also, if the service and Account Owner don't have a previous relationship (e.g. an account at the service), the relationship is assumed to get established during the linking process, e.g. by the Account Owner creating an account to the service.

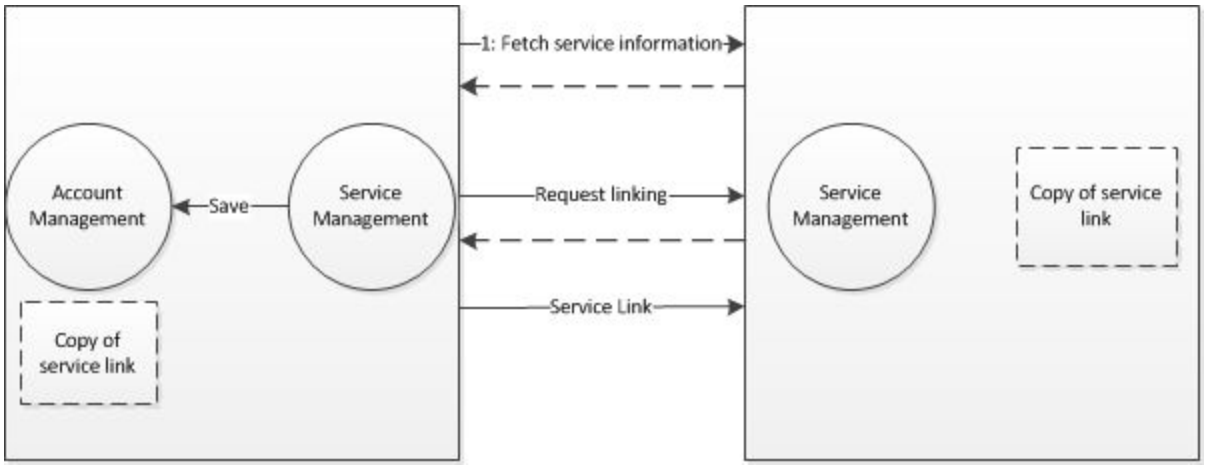


Figure 6.1: The Service Linking process

As shown in Figure 6.1, Service Linking consists of four steps:

1. Operator's Service Management -component fetches information needed to start Service Linking process. Depending on the deployed service registration model this information comes from service registry or from the service.
2. Operator makes service linking request to service. During this step Account Owner must identify and authenticate herself to service, for example by logging into her user account at the service.
3. After user has accepted linking on the service side, the service fills missing parts of the linking record, signs it and sends it to Operator's service management endpoint (SM) for signing. Link record is then signed at the Operator using user's key, saved to user's account and delivered back to service. This completes the linking flow.

During service linking process, a generated surrogate ID is associated with Account Owner's account at the service. This ID is a pseudonym that is meaningful only to Operator and this specific service. It is used in communication between these two parties whenever they need to unambiguously refer to a specific Account Owner's MyData Account (messages from service to Operator), or to a specific user account at the service (messages from Operator to service).

7. Contract-based processing

Contracts are one of the legal bases of processing and a key element in defining the relationship between the individual and a service.

MyData Architecture Framework plans to offer the following contract related functions:

- MyData Account provides digital locker functionality for saving or updating copy of contract or a link to contract residing at the contracted service
- Service can push notifications about contract updates (user then goes to service, accepts new terms, contract copy updated on Operator)
- More advanced functionality is support for creating and negotiating the contract

The current version of the Framework does not define any of the contract-related functions, they are left for future versions of the Framework.

8.Consent-based processing

This chapter covers acquiring and managing digital proofs of Account Owner's consents to Services. A valid consent is documented into a consent record created at point of consent. Consent records also form the enforcement layer for data authorisation that is a necessary technical step for data transfer to a Sink Service in 3rd party re-use use case.

8.1 Acquiring a Consent

Consent needs to be given in a clear manner so that the data controller can demonstrate that a valid consent has been given. This requires maintaining a record of consents and in some cases implementing a comprehensive consent management framework. An effective audit trail of how and when consent was given is needed so that evidence can be provided if consent is challenged.
Consent records should demonstrate the following:

- **Who consented:** the name of the individual, or other identifier (e.g. online username, session ID).
- **When they consented:** a copy of a dated document, or online records that include a timestamp; or, for oral consent, a note of the time and date which was made at the time of the conversation.
- **What they were told at the time**: a master copy of the document or data capture form containing the consent statement in use at that time, along with any separate privacy policy, including version numbers and dates matching the date consent was given. If consent was given orally, records should include a copy of the script used at that time.
- **How they consented:** for written consent, a copy of the relevant document or data capture form. If consent was given online, records should include the data submitted as well as a timestamp to link it to the relevant version of the data capture form. If consent was given orally, a note of this made at the time of the conversation should be kept - it doesn't need to be a full record of the conversation.
- **Whether they have withdrawn consent:** and if so, when.
- **In case of a minor:** Age of a minor and role of the consentee acting on behalf of the minor at time of consent.

8.2 Consenting

There are two types of Consenting supported by the MyData Architecture Framework: 1) consenting to processing within a service and 2) consenting to sharing data from a service (Source) to be processed in another service (Sink). Consenting to processing within service results in a single consent for the service and consenting to 3rd party re-use results in two legally valid consents which are documented in Consent Records associated with the Service Link Records for the service provisioning the data (consent to provision data) and for the service processing the data (consent to process data).

Consenting proves there exists Account Owner's permission for data processing or provisioning. The Account Owner decides in the process, what data a service can process or provision and how the data can be processed and in case of service offering external access to the data, further re-used.

Consenting can happen only after the service has been linked to MyData Account. MyData Account keeps track of all consents a user has given, including expired and withdrawn consents. The user interface of Operator's Account Management Service needs to be able to represent consenting summaries in a clear human-readable form in order to offer an unambiguous description of the consent is in place.

The Account Owner can, at will, deactivate the consent, in which case no new data is processed or provisioned to 3rd parties covered by the consent. If the consent is reactivated, it is service-dependent whether the data collected during the deactivation period now becomes available to the Service (or 3rd party Sink) or not. An example would be Account Owner's current location data, which can be made unavailable for a period and the data for this period will not become available even after reactivation.

Account Owner can withdraw a consent at any time. Authorisation Management Service sets the consent to 'withdrawn'-state and informs the impacted services. Unlike a deactivated consent, a Consent that is withdrawn can not be re-enabled.

Consenting is documented in detail in Consenting Specification available at https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs .

8.2.1 Consenting to processing within a service

For authorising data processing within a service, the Account Owner consents the service to process data within the service under rules and constraints (i.e. policy) set by the Account Owner.

8.2.2 Consenting to 3rd party re-use

For authorising data transfer from a specific Source to a specific Sink for processing, the Account Owner consents the Source to provision the data and the Sink to request and process the data.

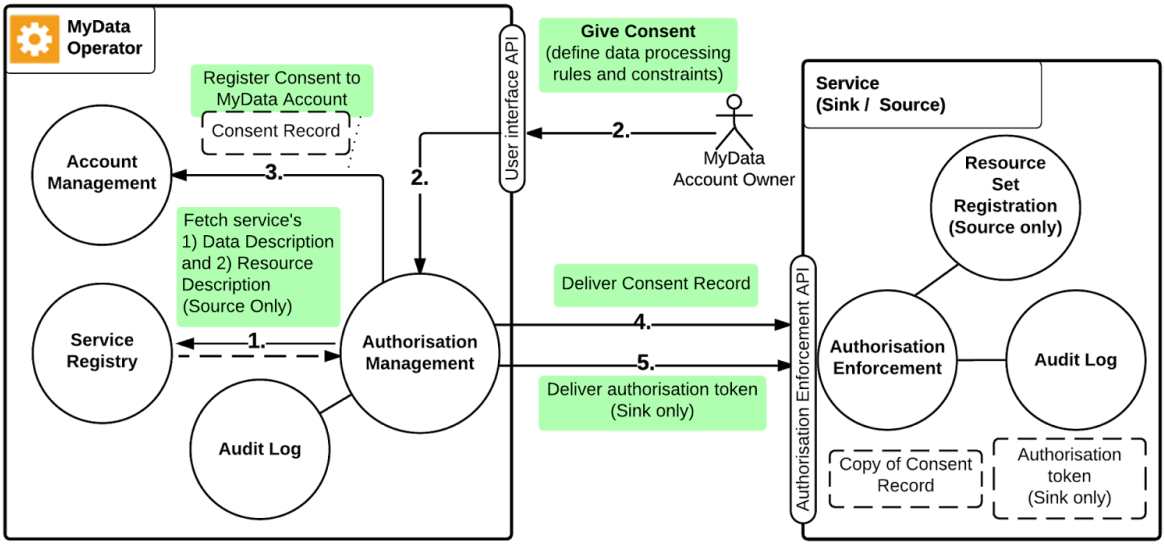


Figure 8.1: MyData Authorisation for 3rd party re-use consist of five steps

Consenting is initiated by Account Owner and it consists of five steps as shown in Figure 8.1:

1. Information about the service is presented to Account Owner including what information the Sink would like to have and for what purpose. Account Owner then defines data processing rules and constraints, which must meet the Sink's minimum requirements for the consent to be actionable.
2. Consent Record is stored in Account Owner's MyData Account.
3. Consent Record is delivered to the Service.
4. Relevant authorisation data (e.g. token) is delivered to the Service

The way Source provisions data described by resource set description, the actual data requests sent by Sink, and further details about token usage are described in 8.3.1 below.

8.3.1 Data transfer to 3rd party

A data transfer is the event where an authorised transfer of MyData Account Owner's data from a Source to a Sink is made. After the Consenting is issued, authorised data transfers may happen from Source to Sink as long as the consent is not deactivated or withdrawn.

Data transfer consists of three steps as shown in Figure 9.1 and covered in detail in [6]:

1. Sink makes a data request to Source using the credentials it received from Operator during consenting process. Sink must verify the related Consent Record is valid before making the request.
2. When Source receives the credential, it first validates the credential, then verifies the Consent Record with which the authorisation token is associated is still valid and active. Source may also make a status check for the Consent Record from the Operator, if e.g. this is particularly high risk data or the Source has reason to believe the status might have changed.

3. Based on the validation, Source either grants or denies the data request. As Sink requested data with only the Resource Set ID, Source uses the Resource Set Description in the Consent Record to determine, what data should actually be given.

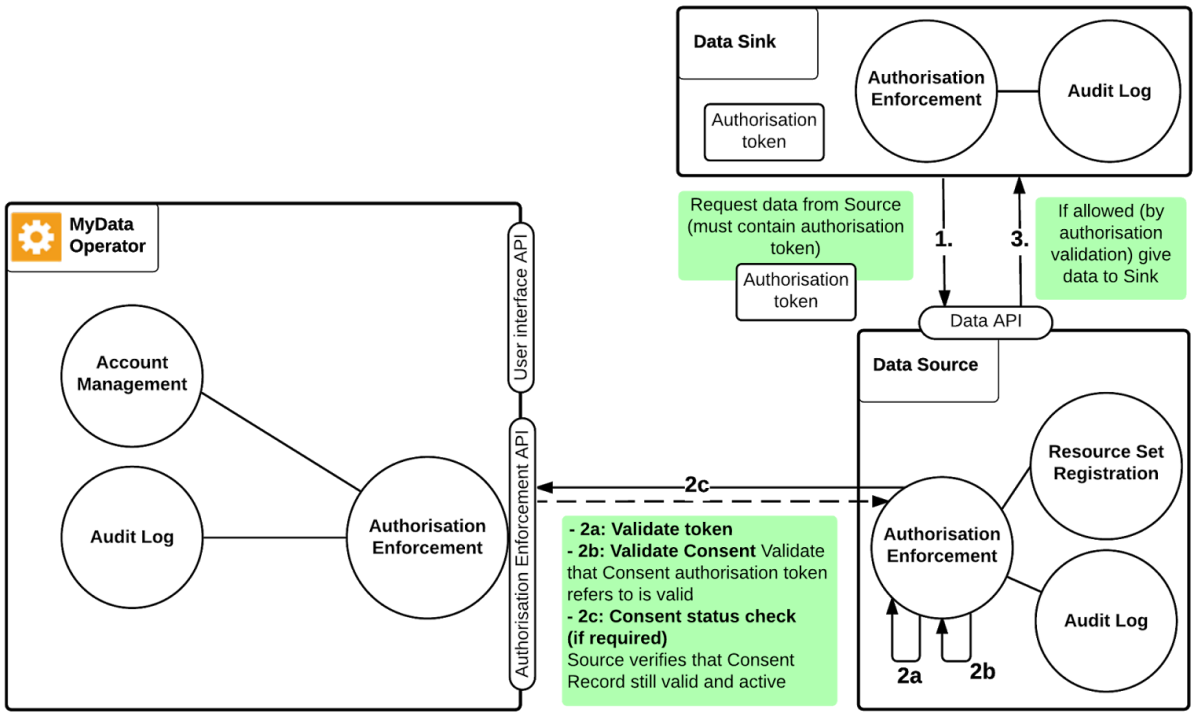


Figure 9.1: Data transfer process

More detailed technical description of the consenting and 3rd party data transfer authorisation process can be found from the technical specifications available at https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs

9. Notification of processing

Notifications enable Account Owner to receive messages about personal data processing. Notifications are subscription based i.e. Account Owner has to make a subscription to receive data processing notifications from a service. Service sends new notification to Account Owner (and operator may have multiple mechanisms supported to bring this information to Account Owner's attention - these are not covered by this Framework) when subscribed event happens.

Possibility to receive processing notifications is dependent on service listing this as a valid processing base in its Service Description.

Notification consists of three steps as shown in Figure 9.1:

1. Account Owner must subscribe to receiving Notifications from the service.
2. Service confirms notification subscription has been activated.
3. Account Owner receives notifications whenever sent by the service.

Notifications may be encrypted and digitally signed depending on the sensitivity of the notification. Serving a notification history as part of the Account Owner's services is recommended but exact notification record format is not enforced by the Framework.

Account Owner can manage notification subscriptions and cancel a subscription at any time.

Notifications are defined in more detail in Notification Specification available at https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs.

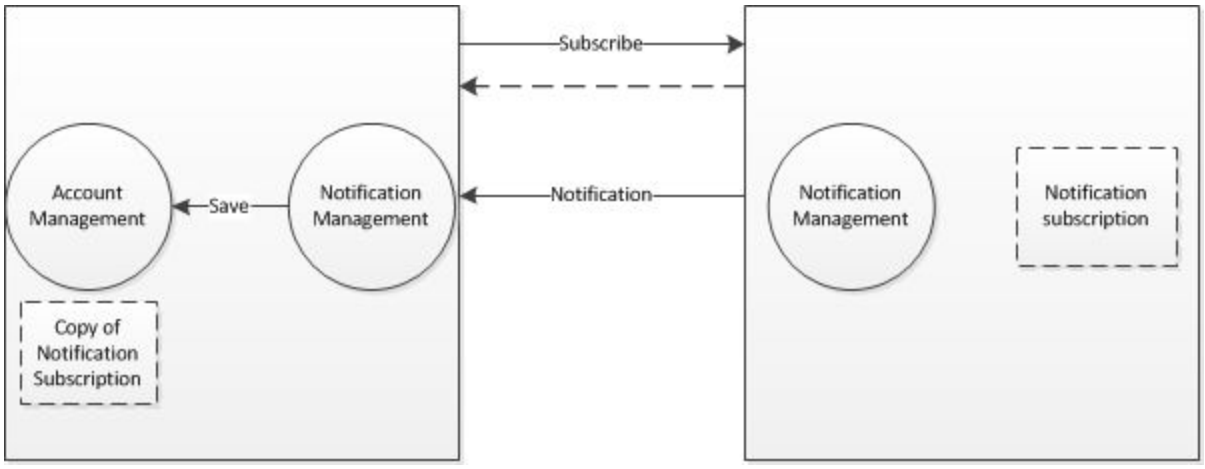


Figure 9.1 Notification process

The current version of the Framework reference implementation does not implement all of the notification-related functions, they are left for future versions of the Framework.

10. Objection to Processing

Objections enable Account Owner to object data processing on a service. Typically availability to object processing is enforced by regulatory bodies regarding certain types of data where processing is allowed by default - there may be national variation on where objections are possible. Possibility to object is dependent on service listing this as a valid processing base in its Service Description.

Objection consists of two steps as shown in Figure 10.1 and covered in detail in Objection Specification available at https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs ..

4. Account Owner must identify and authenticate herself to service.
5. When receiving the Objection request the Service will check the current status of the processing related to Account Owner.
6. If processing is active and Objection is valid function regarding this Service, the processing is set to cease and the Account Owner's status at the service is marked 'Objected' with the details of the Objection request received.

Note that once objected, processing of data can only be activated by offering the Account Owner option to consent to data processing (most likely with original processing terms). Objection is further defined in Objection Specification available at https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs.

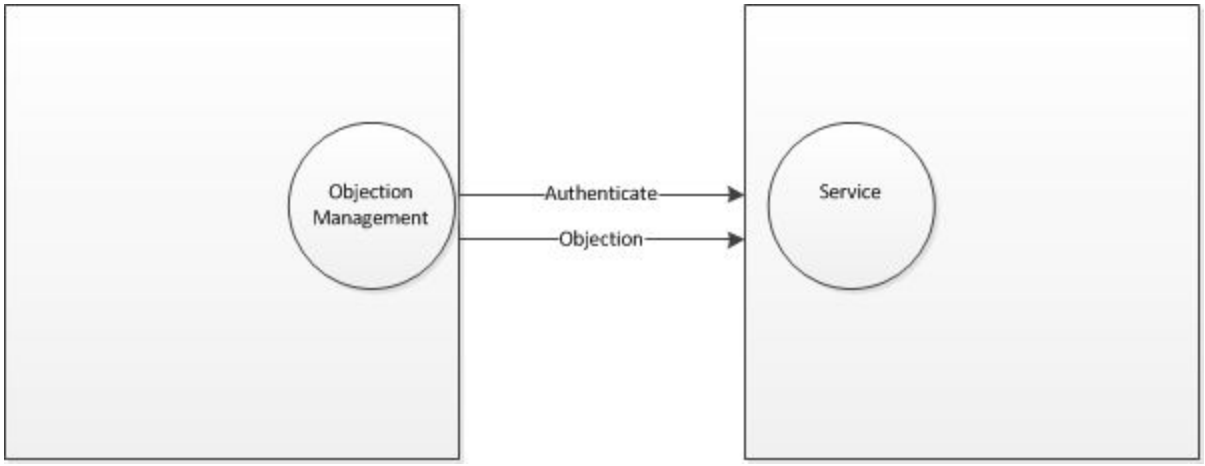


Figure 10.1 Objection transaction

The current version of the Framework reference implementation does not implement all of the objection-related functions, they are left for future versions of the Framework.

11. Compliance

Operators and services are free to choose what use cases they support. Operators MUST support at least MyData Accounts and Service Linking and MUST provide Operator Configuration information. Services MUST support at least Service Linking and MUST provide a Service Description with at least one data and processing description.

References

[1] MyData White paper, http://urn.fi/URN:ISBN:978-952-243-455-5.
[2] MyData Account, https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs
[3] MyData Service Registry, https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs
[4] MyData Service Linking, https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs
[5] MyData Authorisation, https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs
[6] MyData Data Connection, https://github.com/mydata-sdk/mydata-docs/tree/master/api_specs

Appendix 1: Glossary

Account Owner (AO) [role] is the natural person controlling a particular MyData Account. Depending on the account type, the owner may be either (strongly) authenticated or even anonymous. Account Owner is the same as the Data Subject.

Account Management Service is the user interface and related service for managing MyData Accounts, Service Links, Consenting, Contracts, Notifications and Objections.

Consenting [interaction] Account Owner's act of granting permission for 1) a service to process data or 2) data transfer from a specific Source to a specific Sink. 1) results in a Consent Record and 2) results in a pair of Consent Records (one each for the Source and the Sink) documenting the granted permission.

Consent [lawful basis] is one of the grounds for lawfulness of processing personal data and it is given by the Data Subject for one or more specific purposes (Arts 6-7 GDPR). Consent means "any freely given, specific, informed and unambiguous indication of the data subject's wishes" by which agreement to processing of their personal data is signified, either by statement or clear affirmative action.

Consent Record (CR) documents the permission the Account Owner has granted to a specific service. For consenting to data processing within a service, the Account Owner creates a single Consent Record for the related service. For consenting and authorising data transfer from a specific Source to a specific Sink (3rd party re-use), the Account Owner creates a pair of Consent Records (one for the Source and one for the Sink). The Source's CR defines, what data can be provisioned to the specified Sink, and the Sink's CR defines, how the data can be accessed. The Sink's CR can also include the permissions for data processing. A Consent Record is a manifestation of legally valid Consent and makes it technically feasible to change or withdraw the consent dynamically. Consent Records are stored in the MyData Account and with service(s) involved.

Consent Status Record (CSR) is a record MyData Operator sends to a service when status of a consent changes. Service MUST store these records for future use.

Data Transfer is an authorised transfer of data from a specific Source Service to a specific Sink Service.

Data Controller [legal role] is a natural or legal person, public authority, agency or other body which alone or jointly with others determines the purposes and means of the processing of personal data (Art. 4 (7) GDPR).

Data Processor [legal role] is a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the data controller (Art. 4 (8) GDPR).

Data Subject [legal role] is an identified or identifiable natural person whose personal data is processed (Art. 4 (1) GDPR). The data subject has rights and practical means to control creation, flow and usage of his personal data. The data subject gives and manages Consents related to their own data and Service Links.

Data Transfer Log [record] records the Data Connections and unsuccessful attempts of Data Connection. It can be audited using Account Management Service.

MyData is the subset of personal data that the individual can access and control.

MyData Account is a human centric concept in MyData architecture. MyData Account contains Account Owner's digital identity or identities, linked services, contracts, consents, notifications of processing and objections to processing. MyData Account can include additional data about Account Owner to help in providing improved services.

MyData Operator [role]) provides MyData Accounts and the related Account Management Service. Personal Data: any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person (Art 4 (1) GDPR).

Notification [interaction] is the act of a service informing the Account Owner that their personal data is being processed under a legal base other than a contract or a consent to which the Account Owner has explicitly agreed.

Objection [interaction] is the act of the Account Owner requesting a service stop processing their personal data when the processing is not based on a contract or a consent to which the Account Owner has explicitly agreed.

Service Linking [interaction] is the Account Owner's act of linking a service (Service, Source or Sink) to their MyData Account. As the result the Service Linking status and parameters are documented within a digital machine-readable record, called a Service Link Record.

Service Link Record (SLR) is the outcome of a successful Service Linking. It documents in machine readable form the terms and scope of the agreement between the Account Owner and a single Source or Sink. Service Link Records are stored in the MyData Account.

Service Link Status Record (SSR) is a record MyData Operator sends to a service when status of a Service Link changes. Service MUST store these records for future use.

Sink (Service) [role] is an entity that can acquire data from one or more Sources and allows management of data processing through a MyData compliant APIs.

Source (Service) [role] is an entity that can provision data about the Account Owner to one or more Sinks and allows management of data provisioning through MyData compliant APIs.

Surrogate ID is a pseudonym that associates Account Owner's MyData Account to his / her account at the service being linked. This ID is meaningful only to the Operator and the service that generated it. It is used in communication between these two parties whenever they need to unambiguously refer to a

specific Account Owner's MyData Account (messages from service to Operator), or to a specific user account at the service (messages from Operator to service).