MyData Architecture - *The Stack*

Contents

Authors: - Harri Honko - Yki Kortesniemi - Antti Poikola - Antti Kallonen - Samuli Tuoriniemi - Kai Kuikkaniemi - Harri Hirvonsalo - Mika Rautiainen - Ilari Määrälä

Executive Summary

MyData architecture enables users themselves to manage access to their personal data. This document introduces the core concepts, actors, roles and data taxonomies that are necessary in describing the overall architecture of the system.

Currently, applications and services collect increasing amounts of personal data about their users and use it to extract valuable knowledge about them. This knowledge is typically used for profiling users, and the results are monetizable input for e.g. targeted advertising. The users typically have no control over how their data is created or used.

MyData approach proposes a paradigm shift by introducing a human centric model that aims to give control of the produced data back to users. The model is envisioned to e.g. liberate usage of personal health data by letting users to choose how their data can be used by third party organizations. The assumption is that users are the best experts to understand the utility of data about themselves in different applications, and therefore the data will be put to best possible use for the benefit of the user.

In contrast to more relaxed data in e.g. social media services, health data domain is more regulated with higher protection of individual's right to privacy and fair use of data. In addition, the legal framework of the developed societies will protect users from blatant misuse of their health data in situations where the utility and benefits are more difficult to assess. These characteristics make health data a good domain to research and develop MyData system models in various scenarios. However, health is not the only domain benefiting of the model - more casual data, e.g. personal interests, food and media consumption habits, can also be shared according to MyData model to help users have better control over service-specific recommendations, targeted advertisements and other applications that necessitate personal profiling. Users might even be able to monetize their data themselves.

The legal framework for protecting user data starts from the principle of informed and explicit consent. In order for personal data to be available for a third party, the user has to be well informed on how the data will be used and shared and the user has to make a conscious choice to give consent to the external organization to collect data within the limitations of the contract. This is the exact premise upon which this document builds a technical architecture for the MyData model. The implementation of the consent mechanism is inspired by OpenID Connect and User-Managed Access (UMA) specification to authorize sharing and use of personal data resources in online environment.

A key role in the architecture is the *MyData Operator*, an entity providing services to view, manage and control the consents enabling the flow of personal data. Data Operator model makes it possible to execute legal information requests cost effectively and in standardized manner. Data Operator model also empowers users to examine, link, re-use and constrain their personal data in ways that has not been possible with the prevailing paradigm of uneven collection and use of personal data.

Part 1 - Introduction

The simple core idea, ***individual in control of her own data***, is both a political movement for digital human rights and an initiative for opening new business opportunities. Though both of these goals have been possible within the current legal framework, the lack of interoperable implementations has kept them mostly a distant goal - a situation we now want to change.

In this document we present the MyData architecture, 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. It responds on a practical and technical level to individuals growing demand for control over their own digital identity and to organizations need to fulfill the requirements of tightening data protection regulation.

The architecture aims to provide a standard for implementations that

- satisfy the legal requirements for processing of personal data and, thus, removes the risk of sanctions related to improper processing
- enable the users to easily grant and withdraw their consent for data processing
- provide visibility to users about how their data is being used
- enable flexible service creation and new business opportunities

In the following sections we present an overview of the MyData Architecture, provide a technical description of the data flow implementation and detail some future directions for the Architecture.

Part 2 - MyData Architecture overview

This part provides an overview of the MyData architecture. It introduces the core concepts of the architecture, details the legal basis for processing of personal data, describes the main transactions required for data mobility and elaborates on the value adding activities.

Core Concepts

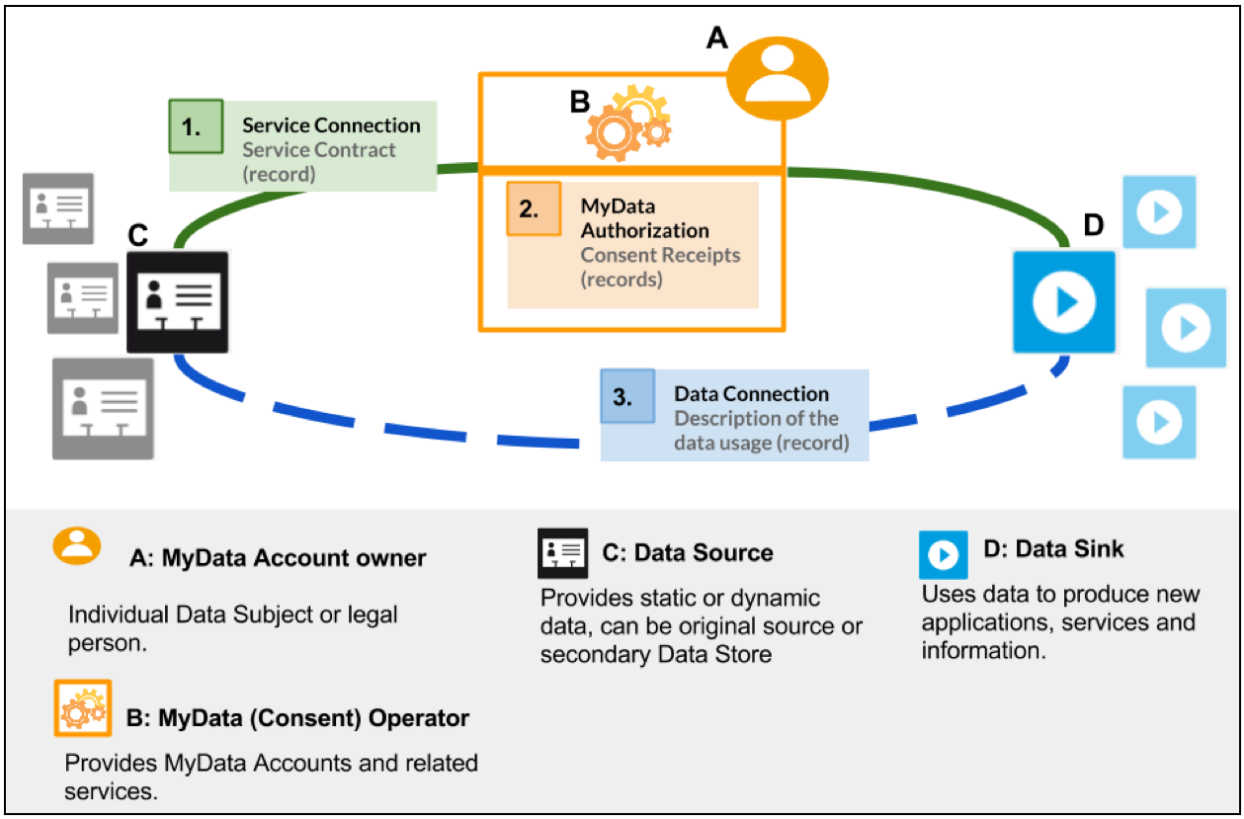

Figure 1. Core concepts of the MyData Architecture

Key concept in MyData architecture is the **MyData Account** which enables individuals to manage the flow of their personal data from many data sources to the services which use the data. There are four operational roles in the My Data architecture: **Account Owner** (usually the Data Subject), **MyData Operator**, **Data Source** and **Data Sink**as illustrated in Figure 1. Actors (organizations and individuals) may work in one or many of the operational roles. It is very typical for example that same organization is in the role of Data Source and at the same time also in the role of data Sink.

We shall next go over each role in more detail.

MyData Account

The account metaphor is familiar from bank, email, and customer accounts. Functionally the MyData Account is the key enabler in authorizing, controlling and logging the data flow between Data Sources and Sinks. MyData Account stores all Service Contracts and Consent Receipts that define access means and rights for Data Sinks and Data Sources as well as the Data Usage Log resulting from authorized Data Connections.

Typically the accounts are provided by organizations acting in the operator role such as teleoperators or banks. In MyData Architecture the operator role is also defined (see Figure 2) and it is expected that most of

the MyData Accounts will be provided by some organizational MyData Operator. However, it is also possible for individuals to run the operator server software themselves and become self-operators, and therefore have functional MyData Account independently of any organization.

MyData Account Owner

Account Owner is the person who created and is using the account to link new services (see Service Connection) and authorize data flow (see MyData Authorization). One person may have multiple accounts, but an accounts may also be shared with multiple people. The Account Owner has control over her account settings.

Depending on the account type and offered security level, the owner may be strongly authenticated, be known by a pseudonym or even be completely anonymous, and the formal registration mechanism of accounts may vary (ex. strong, 2-phase authentication or simple email account/password mechanism). In some domains, such as sensitive health data, the Data Sources may require certain level of security and authentication based the legislative demands. For example, if a person wants to use a MyData Account in a public healthcare service, there might be government regulations that require the use of strong authentication. Also in such a case the public sector healthcare service may require specific certification of the MyData Operator.

MyData Operator

Principal purpose of the MyData Operator is to provide MyData Accounts and the user interface for managing those accounts. Operator also needs to provide the underlying mechanisms for connecting Data Sources and Data Sinks to the account and creating and managing MyData Authorizations.

Operator may also provide specific value-added services such as additional security e.g. by certifying the Actors, local applications for visualizing and monitoring data, data storage and service brokerage. At this point the account specification has not been defined to cover data semantics, intention modeling, profile information, data anonymization or answer services (such as OpenPDS). In future the MyData Account specification may be extended to cover also such features. The components of the operator software stack are defined in section Operator Stack.

Data Sources and Data Sinks

Data Source is an entity that can provision data about the Account Owner to one or more Data Sinks and correspondingly Data Sink is an entity that can capture data from one or more Data Sources and uses this data to produce new applications, services and information. Both Data Sources and Sinks need to provide a MyData compliant API. Data Source API allows management of data provision, while the Sink API allows management of data usage. It is quite common that same service is both working as Data Source and Data Sink and providing, therefore, both Sink and Source capabilities.

Data Source and Data Sinks are general high level operational roles. There are special types of Data Sources such as a proxy, a data logger, a data store and an MyData Account as a Data Source. There are also special types of Data Sinks such as applications hosted by the individual, physical devices, aggregator (aggregation over multiple identities), data synthesis (aggregation over multiple sources of same individual) and anonymizer to mention few.

Legal basis

A central question in processing personal data is who decides when and how an individual's personal data is processed: that person or others? According to international regulation of data protection, such as the EU Data Protection Directive (95/46/EC of 24 October 1995) or the Charter of Fundamental Rights (2010/C 83/02), there are several legal bases for processing personal information. For example, others may process personal data on the basis of contracts, legal obligation, etc. without consent from the individual.

However, at the heart of the rights of the individual, 'the Data Subject,' is that personal data is processed on the basis of that individual's explicit consent. Consent is also the very foundation the MyData architecture approach: data processing is always based on the user's valid consent and no processing takes place without valid consent. For a consent to be valid, is has to fulfil a number of legal requirements - and in EU these requirements are about to change due to the upcoming EU General Data Protection Regulation, the GDPR (NOTE: The new EU-wide regulation will be based on Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data.).

The MyData Account based system architecture incorporates the increasing legal requirements regarding consent, the rights of the data subject, and other demands relating to processing as being enacted in the GDPR.

It should be noted that GDPR is not yet in force and still being amended. Thus, as of spring 2015, the final legal obligations relating to data processing are not yet known. However, their development and changes are being closely followed and if needed this MyData specification will be adjusted to the changes.

Consent

Consent is the permission to process personal data and it is always given by the Data Subject to the Data Controller. To be legally valid, the Data Subject needs to give an unambiguous consent to the processing of their data to *specific purposes*, and the consent has to be *free*, *informed* and *explicit*. Consents can be specifically created or part of a larger (not just MyData-context) legal contract between the individual and services she uses.

In MyData architecture consent is a central concept as all transactions are based on consents (though there exist other legal bases for processing of personal data, in the MyData Architecture all processing requires a consent from the Data Subject). It allows Data Subjects to decide how their personal data are being used and make decisions about how they can be used in the future. This approach is also geared towards the requirement that withdrawal of consent has to be as easy and swift as giving it. The MyData Account infrastructure makes it technically feasible to change or withdraw the consent as needed. This applies to all types of personal data, irrespective of whether they belong to special categories of data known as sensitive personal data (NOTE: Personal information revealing race or ethnic origin, political opinions, religion or philosophical beliefs, sexual orientation/gender identity, trade-union membership and activities, the processing of genetic or biometric data, information concerning health or sex life, administrative sanctions, judgments, criminal or suspected offences, convictions or related security measures.).

Legal and Operational Roles

In the data protection regulation there are a number of legal roles, most importantly the Data Subject, Data Controller and Data Processor which have differing rights and obligations. Data Subject is an identified or identifiable natural person *whose personal data is processed*. Data Controllers and Data Processors are either natural persons or legal persons, public authorities, agencies or other bodies. Data Controller alone or jointly with others *determines the purposes and means of the processing of personal data* and the Data Processor processes personal data *on behalf of* the Data Controller. Data Processor role is not defined in U.S. privacy legislation so far.

For unambiguity it is important to make mapping between legal roles - Data Subject, Data Controller and Processor - and previously presented operational roles MyData Account Owner, Data Source and Data Sink. It is worth noting that the legal framework only covers cases of personal data where the Data Subject is natural person, but the MyData architecture technically works equally for managing data of juridical persons or even anonymous Account Owners.

When the Account Owner is natural person and her data is treated she is allways the Data Subject. The other legal roles in MyData system can be then determined by answering two questions:

1. Who determines the purposes and means of the processing of personal data?
2. Who actually does the data processing?

Some typical cases are:

- **Repurposing:** The Data Source is also Data Sink processing personal data for a specified purpose - at some point they may suggest for the Account Owner a new purpose or means of processing data. In this case the Data Source is in legal terms the Data Controller.
- **Delegation:** The Data Sink accesses with Account Owners consent personal data from the original Data Source and processes it for a defined purpose. In this case both the Data Source and the Data Sink are in legal terms Data Controllers.

Beside these typical cases the MyData architecture finally makes technically feasible the long-anticipated concept, where the Data Subject can ultimately also act as her own Data Controller. The upcoming EU General Data Protection Regulation describes in its published terms a means to implement scenarios where an individual can act as Data Controller for her self-managed secondary data usage rules & terms over her self-managed personal data that can be re-used, and consents for these secondary or re-uses of the data are managed by the person herself. It states that an external actor (such as a service provider) can offer services related to this data without implicitly becoming a Data Controller. Of course in several cases the external actor will have internal interests or regulatory responsibilities that will trigger it also to be a Data Controller.

Novel case:

- **Assignment:** The Account Owner defines the purpose and means of data processing, but processing is done by a separate Data Sink which accesses data from the original Data Source. In this case the the Data Source and Account Owner are both in legal terms Data Controllers and the Data Sink is a Data Processor processing the data on behalf of the Account Owner.

Transactions

At the highest level there are there are three main transaction types between Data Sources, Data Sinks and MyData Operator:

1. **Service Connection:** linking a new Data Source or Data Sink to a MyData Account
2. **MyData Authorization:** authorizing specific Data Source to provide data for a specific Data Sink and the sink to use that data.
3. **Data Connection:** establishing authorized automatic data transfer from a Data Source to a Data Sink

All these transaction types result in digital records that are stored at the MyData Account:

1. **Service Contract** is a record of an established Service Connection
2. A MyData authorization results in a pair of **Consent Receipts**, one each for the Data Source and the Data Sink
3. All data connections are recorded in a **Data Usage Log**

Service Connection

Service Connection is an action executed by an Account owner to link a service (Data Source or Data Sink) to her MyData Account. As the result the Service Connection status and parameters are documented within a digital machine-readable record, called a Service Contract. One legal party of the Service Contract is always a Account owner. This digital contract defines how the Individual uses the service and whether connected service is a Data Source, a Data Sink or both. Contract's machine-readable internal structure defines input and output data sets, service API endpoint, and agreed rules on how the data is to be handled.

MyData Authorization

MyData authorization proves there exists Account owner's permission for active data transfer from a specific Source to a specific Sink. Authorization results in two legally valid consents which are documented in Consent Receipts attached to the Service Contracts with the Data Source (consent to give data out) and Data Sink (consent to use the data).

The Consent Receipts (as well as the Service Contracts) are machine-readable digital records stored at the MyData Account. In order to meet the legal framework requirements i.e. to offer an unambiguous description of the details of the consent the user interface of the Account Management Service needs to represent consent summaries in clear human-readable form. The creation of authorizations follows the Privacy by Design principle and are, therefore, always as restrictive as possible.

Data Connection

A Data Connection is the event where an authorized transfer of MyData Account owner's data from Data Source to Data Sink is made. After a MyData Authorization is created, multiple data connection events may happen from data source to data sing as long as the authorization is not deactivated or withdrawn. In order for a Data Connection to happen, a Data Sink must first make a request-for-data to the Data Source, providing proof-of-authorization (proof that this Data Sink is authorized to fetch data from a Data Source) with this request.

Occurrence of Data Connections between Data Source and Data Sink are recorded in a Data Usage Log. This log can be can be used for auditing purposes. E.g. the current databases used in hospitals automatically log each time when someone accesses medical records so that the it is possible to later catch any unauthorized use.

Operator Stack

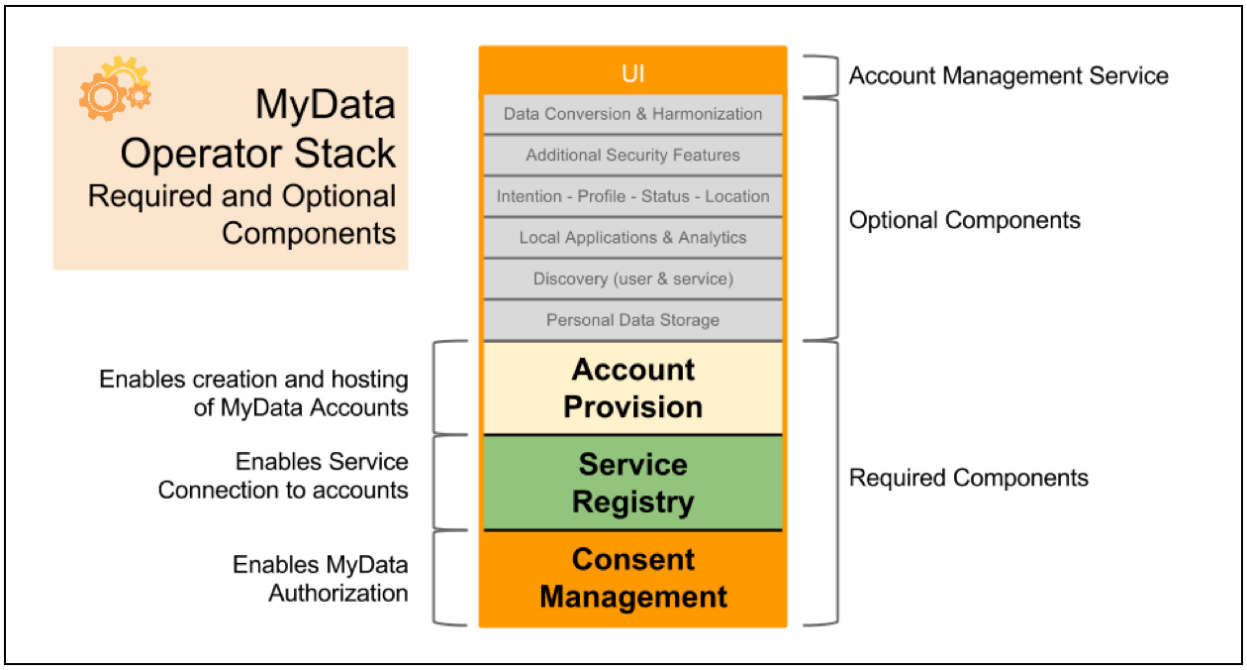

Figure 2: The MyData Operator Stack shows the required and optional functionalities of a MyData Operator

Part 3 - Technical Implementation of MyData Architecture

This section presents the technical implementation of the MyData architecture. For this version of the document we have decided to focus on describing a data flow model, which covers steps needed to establish a Data Connection between Data Sink and Data Source. The presentation follows the same three main steps defined in the Transactions section: Service Connection, MyData Authorization and Data Connection. Future versions of this architecture document will present additional technical details (e.g. data taxonomies and the registration of Data Source and Data Sink to MyData Operator).

Data flow model

Implementation of the data flow model in the MyData architecture is based on the

UMA defines an implementation for a data connection between two (previously unrelated) services (one providing user's personal data, one wishing to have access to the data) so that a human user can securely manage the access to her data. MyData Architecture follows the UMA model and APIs whenever suitable for our approach. However, as MyData Architecture introduces the Data Operator role that helps introduce the Sources and Sinks to each other, we are able to simplify the authorization process compared to what was originally proposed by Kantara.

Resource Sets

A key concept for the data flow model is Resource Set: it defines a specific (subset) of the Data Subject's data that a particular Data Source provisions. This means that the same data a Data Source has can be provisioned in different ways by defining different Resource Sets. In MyData Architecture the Data Source contacts the MyData Operator using the Protection API and registers the Resource Set. After this, the Resource Set can be used in Mydata authorization and each authorization then refers to a particular Resource Set.

Interfaces of different actors

There are four main interfaces in MyData Architecture. MyData Operator provides a *Protection API* for the Data Source. This API defines methods for Data Sink to 1) register resource sets to MyData Operator and 2) introspect received data access tokens from Data Sink. Protection API is protected by OAuth 2.0-protocol and thus each message Data Source sends to this API, must contain an access token, which in this case is called a Protection API Token (PAT).

In a similar fashion, MyData Operator provides an *Authorization API* for the Data Sink. This API defines methods for Data Sink to receive 1) Requesting Party Tokens (RPT) and 2) Consent Receipts in order to access data provided by Data Sources. Each message Data Sink sends to this API, must contain an Authorization API Token (AAT).

In order for Data Sink and Data Source to come into possession of their relevant access tokens, they must provide OAuth 2.0 defined Authorization Grant and "exchange" it for an access token through their respective APIs - Protection API for Data Source and Authorization API for Data Sink.

Both the Data Source and Data Sink must provide an API (*Data Source API* and *Data Sink API*, respectively) for MyData Operator, through which MyData Operator can deliver Consent Receipts, RPT tokens (only for Data Sink) and notifications about changes in consent.

Before Service connection, MyData authorization and Data Connection phases can happen all of the communicating parties must be known. From technical perspective, this means that when connecting Data Source to Data Sink with MyData Operator, these services must know which API endpoints of each other they must call on the network level and what are their identities. Relating these service endpoints to

communicating parties can be performed either at service level when registering Data Source and Data Sink services to MyData Operator or they can be discovered from dynamic service linking URI:s. Choice of endpoint address delivery depends on operator implementation and required security level.

Unless stated otherwise all interfaces are HTTP based RESTful APIs. Structure of security tokens (RPT, PAT, AAT, etc.) used in this data flow model is JSON Web Token (JWT).

Service Connection

The first step for a MyData Account Owner that wants to manage access to his/her data on a Data Source through MyData Operator, is to make Service Connection with this Data Source. In a similar fashion, the Account Owner must make a Service Connection with a Data Sink, but in this case the Service Connection specifies what kind of data a particular Data Sink is to access and how a particular Sink is to use this data.

The creation of a Service Connection requires that the Data Sources and Sinks provide a MyData Service Contract Template (SCT), and publish their template (Table 1 presents the tentative contents of a SCT) at MyData Operator's service registry. The exact process of how this existence is formed between the Operator and actors is not specified in this version of the Architecture document - it is a service registration process that can require human interactions between the Actors, or Operator's own approach to automate any service registration and thus SCT acquisition within its domain.

Table 1: Tentative structure for a Service Contract template (SCT), all fields will be defined in later versions of this architecture document.

```
---
actor_id: {}             # GUID used to identify the MyData actor globally
endpoint_uri: {}         # for the source or sink API

scopes:
    role: {}                # default:source / sink / both
    legal_role: {}          # default:controller / processor
    contract_terms: {}      # default:none / URI (may be legacy or MyData consent)
    intended_use: {}        # default:free / comm-sell / comm-keep / anon-research /…
    usage_rule_set:
        validity_period: {}     # default:auto_renew / valid_until
        auth_proposal: {}       # default:for_data_sink / experimental / standard

user_id: {}              # default:not_defined (Local identifier of the Account Owner
                         # at the sink or source service. Acquired M2M from each service
                         # only at the registration phase.

status: {}               # default:not_defined / active / passive / revoked / void
created: {}              # UTC_timestamp
audit_log: {}            # UTC_timestamp + log_content (vocabulary to be specified)
```

A Service Contract is created by filling in all the required information and signing it with the Account Owner's and Actor's keys. Signing binds the document to a specific Account Owner and the record is stored in the MyData Account.

It is assumed that the Account Owner making a Service Connection with a Data Source, has an account and has or will have data stored in the Data Source. The data is provisioned as Resource Sets, which are identifiers for a specific set of data. Data Sink's Consent Receipt (see Table 2 for the contents of a Consent Receipt) us a Resource Set to define what resources residing on the Data Source, Data Sink is allowed to access.

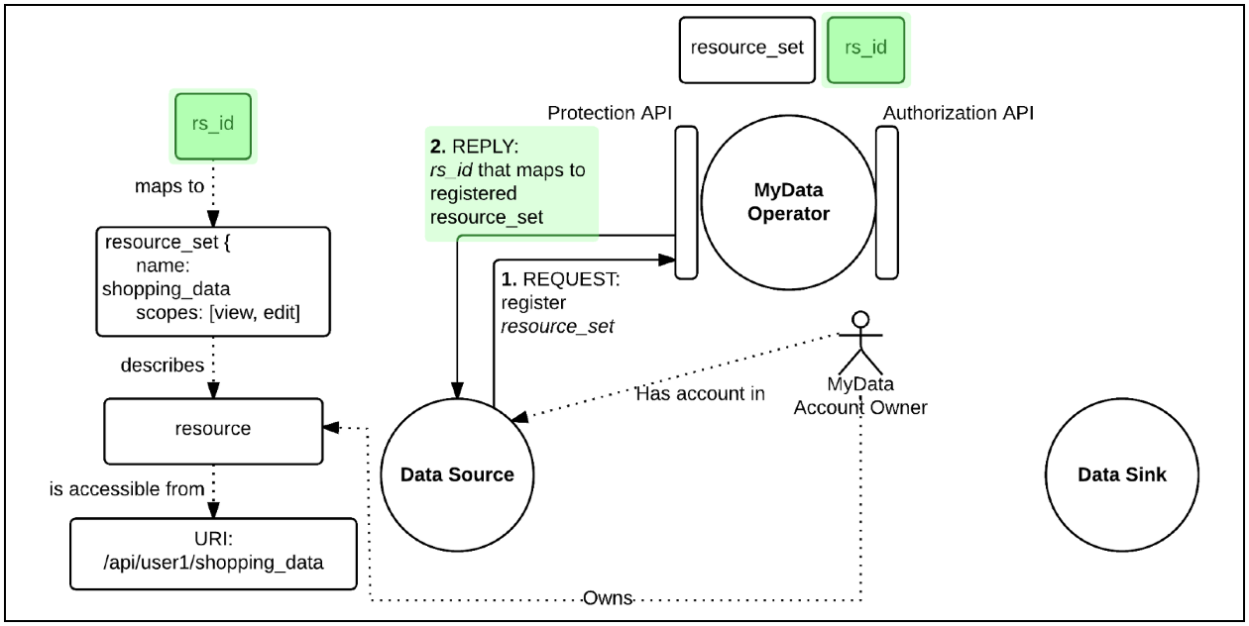

Figure 3: Service Connection - resource registration process that happens after a Service Connection has been made

Forming a Service Connection with a Data Source, triggers a resource registration process, which is presented in Figure 3. In the resource registration process, Data Source begins registering Account Owner's resources to MyData Operator, presenting them as the aforementioned Resource Sets (1). Upon receiving this resource registration request MyData Operator stores the received Resource Set, generates a resource set id (rs_id) that maps to this specific resource set (on this specific Data Source) and sends this *rs_id* as a reply to Data Source (2). Data Source associates the received rs_id with the resource set that it requested the registration for. Resource registration process ends after all the resources have been registered. New Resource Sets can later be defined, if e.g. new types of data becomes available or if the Account Owner wishes to define a new subset of data to provision.

Resource set identifier (rs_id) is composed of a resource locator URI that identifies the Data Source and resource key (a nonce) that identifies the resource inside the specific Data Source. URI and the resource key must not leak out any specific information about the user or data but only provide reference to the Data Source e.g. *datasource.fi/resource/a3h413h4b13h41*. By using rs_id's, both parties can refer to a specific resource set and Data Source using the rs_id as an globally unique identifier. This in turn enables an easy-to-utilize way to implement, for example, the construction of Consent Receipts that explicitly identify what data Account Owner has allowed a Data Sink to access.

MyData Authorization

After a Service Connection with a Data Source has been made, a Data Sink can be authorized to access data on that Data Source by conducting MyData Authorization step. The Authorization results in a pair of Consent Receipts. In the first, Account Owner gives an explicit consent for particular Data Sink to access a specific Resource Set on that particular Data Source (depicted in the Data Source's Consent Receipt). In the second, Account Owner also authorizes Data Sink to process the specified data according terms defined in the Data Sink's Consent Receipt. The contents of a Consent Receipt are depicted in Table 2. So both Data Source and Data Sink have their own Consent Receipt, which contain role specific information necessary in establishing a Data Connection between Data Source and Data Sink. URI that points to Data Source's API endpoint, where the data defined in the Consent Receipt can be accessed, is one example of such information.

Table 2: Data elements in a Consent Receipt (CR)

Receipt ID (unique local ID for the CR)
Account ID (the Account owner's identity)

Information present in Data Source's Receipt:

- *Source SCT*
- *Resource set id (identifies Account Owner's resource / data on Data Source)*
- *What key will be used to digitally sign the proof-of-authorization (an RPT-token) that is given to a Data Sink.*

Information present in Data Sink's Receipt:

- *Sink SCT*
- *What data is Sink allowed to access.*
- *How Sink can use data that it receives.*

Digital signature
Authorization status (active/paused/withdrawn)
Status Change Log (e.g., a latest-first ordered list of: (timestamp for status code change, status code))
Human Consent Summary (human-readable text of the authorized parties and data, can be built based on the SCT and resource set information obtained from Data Source.)

Upon creation of Consent Receipts, MyData Operator delivers each Receipts to their relevant relevant receivers as shown in Figure 4. Data Sink's Receipt is delivered to Data Sink and Data Source's Receipt is delivered to Data Source. In addition, MyData Operator generates a Requesting Party Token (RPT), which is delivered to the Data Sink. This RPT token serves as a proof-of-authorization, which Data Sink uses when making data requests to the Data Source. RPT is contained in each request the Data Sink makes towards Data Source.

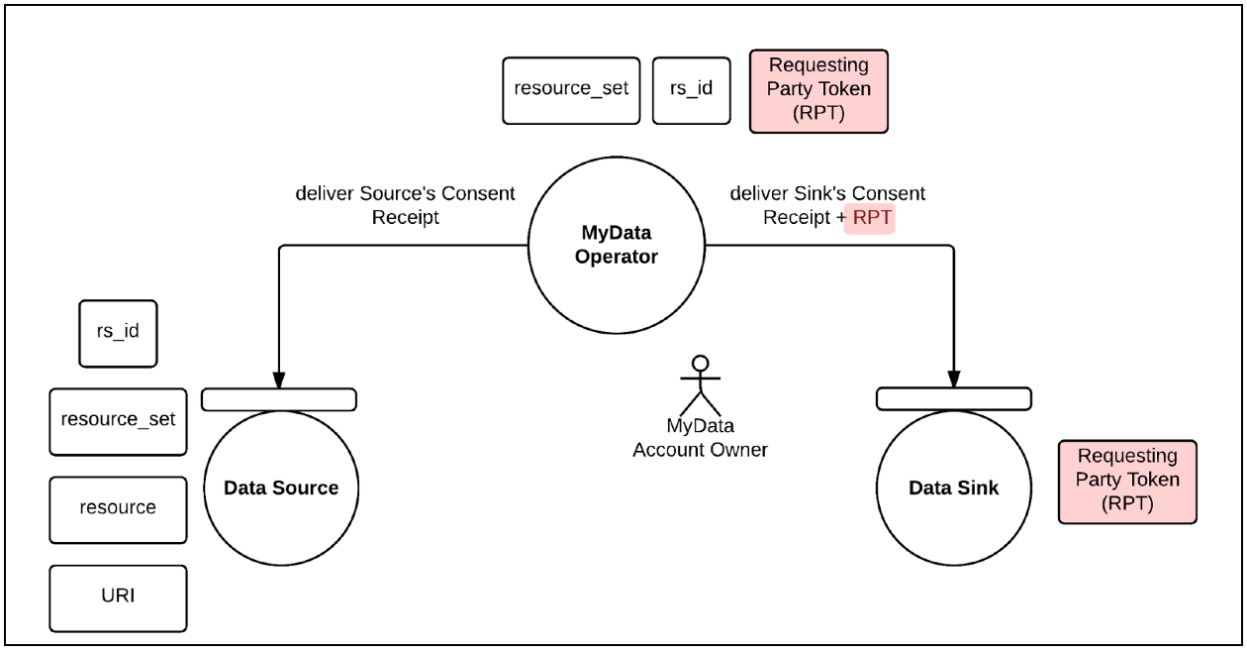

Figure 4: MyData Authorization

Data Connection

Third step of MyData architecture data flow model is Data Connection, where Data Sink authorized to fetch data makes a request containing a proof-of-authorization to Data Source that stores the data. Data Source receiving such request, must check the validity and integrity of the provided proof-of-authorization. Validity check is needed in order to check that user hasn't withdrawn or paused the given authorization, e.g. the authorization is still active.

The actual data flow from Data Source to Data Sink is presented in Figure 5. In the first step Data Sink makes a request to Data Source to access specific data (1). This request contains the MyData Operator generated Requesting Party Token (RPT) that is an access token which serves as proof-of-authorization for the Data Sink service. RPT relates to a Resource Set in the Data Source.

After receiving the RPT, Data Source verifies the integrity of it, and checks that the token is (still) valid, by making a request (know as *introspection*) to Data Operator. (1.1). Operators response contains confirmation that the token is active and what are the current data modify operations for the resource set Data Sink requested (1.2). Finally Data Source delivers the resource set referred by RPT to the requesting Data Sink (2).

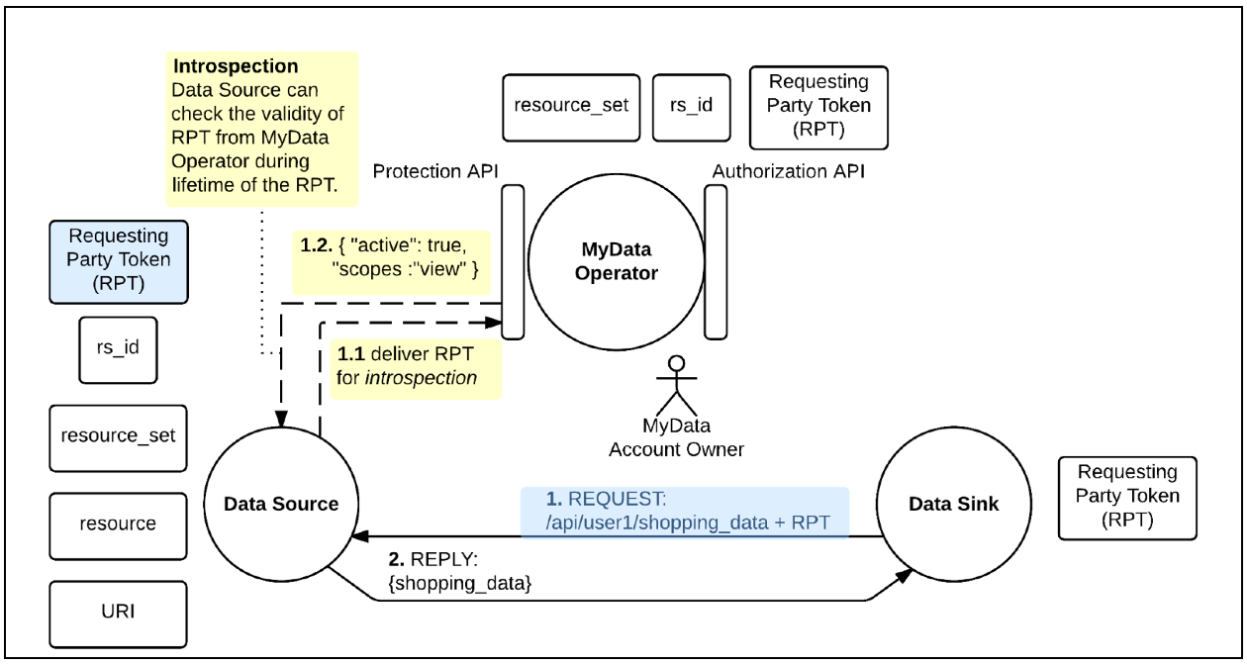

Figure 5: Data Connection

After the initial introspection, Data Source has acquired information that identifies what specific data is allowed to be given in responses to requests of specific Data Sink, and as such Data Source can cache and map this information to the RPT in question. This caching of authorization information in combination with checking the integrity of MyData Operator made digital signature contained within the RPT, enables Data Source to validate RPT without introspection. Thus the introspection steps (1.1 and 1.2) can be performed only periodically as required e.g. for validation. As introspection presumably takes considerable amount more time compared to verifying the digital signature of the token, this time-to-time introspection enables more efficient data transfer between Data Source and Data Sink.

Considering both validation checking mechanisms, RPT's expiration time also has to be taken into account, as RPT's validity can only be verified if the expiration time of the token hasn't exceeded. In other words, no validity checking is done for a received RPT that contains an expired expiration time and the token is rejected immediately. If an RPT's expiration time is exceeded, Data Sink must request a new RPT from MyData Operator. Steps to request and acquire the new RPT are not presented in the diagram.

All Data Sinks have to log all Data Connections (even attempts) for auditing purposes.

Changes to a Consent

Account Owner might make changes to or withdraw the whole consent given to a Data Sink at any time during the lifetime of this Data Sinks Service Connection, at which point Data Source should be informed about these changes as soon as possible. Figure 6 presents an API that Data Source provides for these consent change notifications. Upon receiving such notification from the MyData Operator, Data Source must immediately act and change its processing of request made by Data Sink, according to information presented in the notification. For example, if Account Owner withdraws a consent, MyData Operator notifies Data Source about the change and Data Source starts to deny / reject requests that concern data defined in the withdrawn consent. If a Data Source is processing request at the time it receives a consent change notification, it checks if the information presented in the notification concerns the request-under-processing and updates its request processing if needed.

A corresponding Sink API is used to notify the Sink about the change (this is not depicted in the Figure).

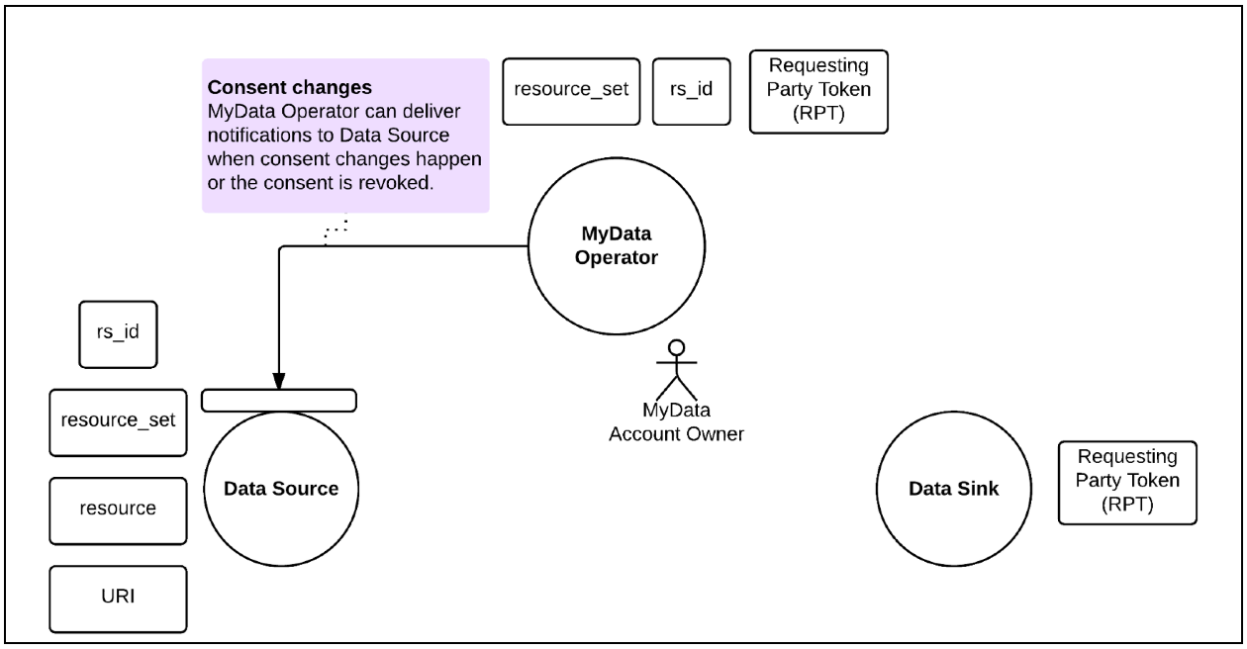

Figure 6: Consent changes and withdrawal

Communication encryption mechanisms

Figure 7 shows the secured communication channels between MyData services. All data transfers between Data Source, Data Sink and MyData Operator are secured with TLS, which encapsulates the HTTP-REST payload protocol. In order to use TLS, these communicating parties must have an X.509 certificate that uniquely identifies services who are transferring data. For development purposes this certificate can be self-signed but for deployment environments it needs to be certified by a trusted certificate authority. When establishing the secure channel the communicating parties must check the signature, expiry date and the revocation status of all certificates in the other party's certificate chain and reject if check fails. All the entities within single MyData Operator domain must have access to same trusted and synchronized time source. Each entity should try to maintain the time and synchronize their clock only when needed. Communication between parties must not proceed if either of the communicating parties notices that their clocks are out of sync. In this case parties should synchronize their clocks with trusted time source before re-establishing the connection. Audit logging of consent operations is performed at the MyData Operator.

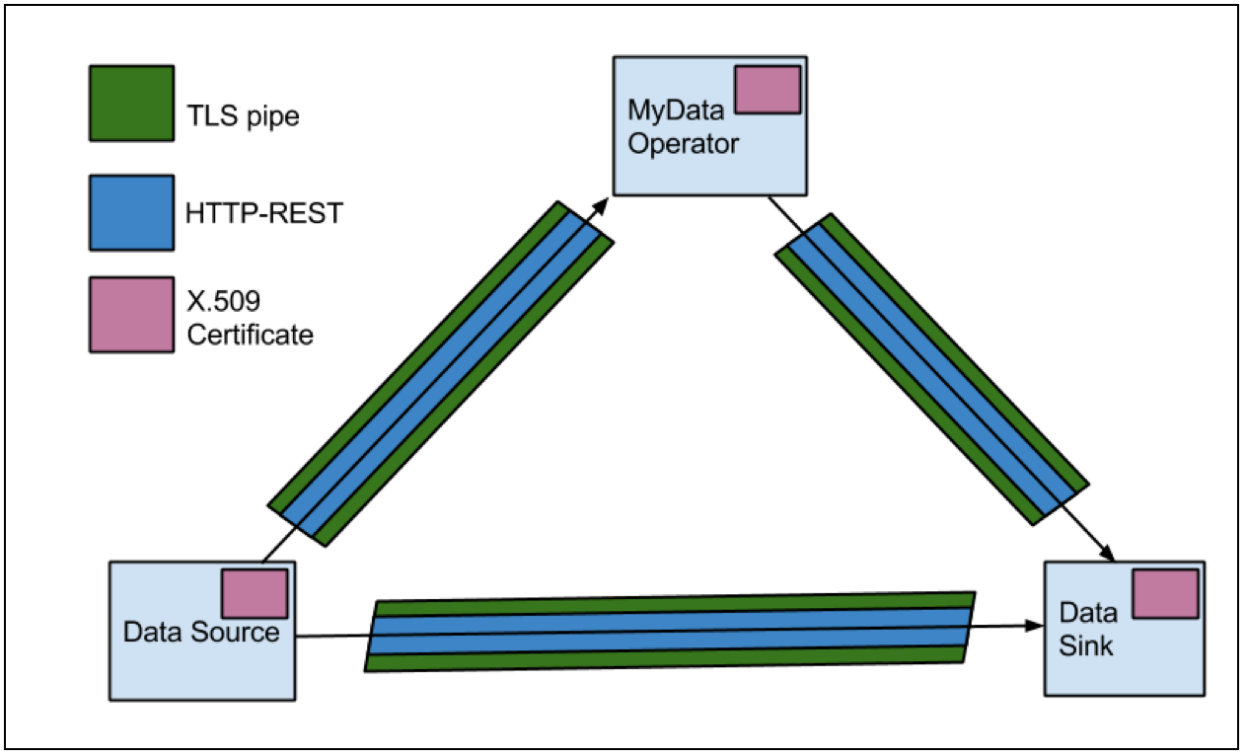

Figure 7: Secured communication channels between MyData services.

Actor Revocation

Each MyData actor has a X.509 certificate that uniquely identifies the actor. Certificate's validity is verified every time it is used using OCSP [or other suitable mechanism] and only valid certificates can be used for establishing SAC between actors. Certificate's issuer [or party acting on behalf of the issuer] can revoke actors certificate i.e. make it no longer trusted. If the certificate is revoked the actor can not any longer participate in MyData transactions. Actor revocation can happen for example due to compromised private key or due to policy violation. The actual reasons and policies for actor revocation are business decisions and are not specified in this document.

Multiple operators

The architecture enables the existence and use of multiple operators. Each user can choose to use one or more operators to manage their consents.

If user decides to change MyData Operator to some other data operator implementation, this requires notifying the related Data Sources and Data Sinks that the tokens have to fetch new tokens from newly appointed Operator.

Depending on operator and user security requirements the storage for cryptographic keys used to sign the user-given consents can be located either at the Data Operator servers, or for more secure setups, they can be stored at the user's possession on a personal computer or a smart card.

Part 4 Development Ideas of further MyData Enablers

Part 4 describes new ideas that are worth consideration and development effort in the future. The ideas have been identified during the design of MyData Architecture and they can act as possible enablers and catalysts for MyData adoption "in the wild". The ideas are categorized into three areas: *MyData Consortium and other supporting actors, MyData Operator Value Adding Services* and *MyData Standards and Specifications.*

MyData Consortium and other supporting actors

Important, supporting actors in the MyData ecosystem are different technical providers, core standards organizations (MyData Consortium), legislative bodies and societal actors. They enable and guarantee the functionality of the ecosystem and facilitate the adoption of MyData, but in the described architecture they have no roles.

MyData Consortium

For the future development of the MyData related core standards, clear and transparent ownership and governance needs to be defined. Initial suggestion is to establish a consortium, where MyData operators and possibly other stakeholder organizations are members. Consortium's main focus is to establish the way of working and maintain interoperability between different actors within the ecosystem. This involves technical standardization of the accounts, consent system, addressing system, authorization principles, aggregation principles and, e.g., anonymization practices, certification of actors to different roles, definition and establishment of APIs and semantic models as well as data interchange and storage practices.

Legislative bodies and societal actors

Legislative bodies can enforce the use of certain practices and facilitate that public organization start to provide the required interfaces and consent systems. For some types of data there might be requirements that operators and aggregators are somehow licensed and audited prior to allowing them to operate, and re-audited regularly during operations. On a societal level there is multitude of actions to empower different stakeholders to function properly and utilize MyData, include monitoring, training, campaign against misbehavior and general concept literacy improvement.

Authentication and Reputation Services

Actors can take multiple roles, and identity of each role they take can be authenticated by an external authentication service. Also data can be authenticated, which is relevant in the case if, for example, Account owner is provisioning sensitive data that is originally created and sourced by some other actor. If Data Sink wants to verify the data authenticity, the sink actor can request the source to verify this data by using a separate authentication service. Authentication service can be based on certificates, but at a lower trust level, there can also be reputation-based authentication of the identity.

MyData Operator Value Adding Services and Business Models

These ideas aim at evolving MyData Operators' core capabilities in directions that further facilitate MyData use and utilization "in the wild".

MyData Operator Business Models

Account service may be subject to account fees. Operator may charge an account fee from the account Owner. Operators may be also collecting transfer fees from each other if another operator is managing the organization's end point of the consent and agreement management. Such operator billing federation is not covered in the current MyData specification. Neither is the current specification considering the monetary logistics of data sales. These are also subject to further investigation and the current specification may be extended to cover these items.

Universal MyData Account Register

In order to be able to always resolve the account(s) used by the user, MyData Consortium would keep a globally directory of the MyData Account IDs (unique name and address of the Account) across the whole MyData system.

Account Mobility

Process that helps the user switch from one MyData Operator to another by facilitating the transfer of Service Contracts and Consent Receipts between Operators - i.e. a person wants to keep her service contracts and authorizations but changes the MyData Operator.

Data discovery

Discovering data resources is key in identifying and linking related Data Sources with Data Sinks. It also facilitates developers to identify possible data sources for their applications. Data discovery is dependent on proper data resource description scheme, which is realized typically by service descriptions in the service registry. In order to provide capabilities for machine to machine exchange of information, structured representations of service resources are necessary. Data taxonomies will be very useful for this task. In particular, many existing health related ontologies provide established representations that can be reused for MyData related services.

The management of data collected from diverse sources is challenging as e.g. the nature, purpose of use or degree of sensitivity may be ambiguous. As envisioned in the scenario presented in Automated Calorie Counter scenario, the purpose of data taxonomies is to support the discovery and utilization of data providing services. In more detail, by classifying different data sources and creating appropriate metadata mechanisms, applications are able to better deduce the services that provide the data they need. In this way, users are liberated from manually searching the right data sources for their applications. Moreover, application developers can be notified as new services providing appropriate data for their applications are registered to the system.

User Discovery Service

One application are for MyData Architecture is for user's providing their data for research purposes either as identifiable information or as anonymised data. A user discovery service would provide the possibility for research institutes and companies to discover users that match certain specified profiles and make requests to use their data. This is based on users having given consent that their profile information can show up in the searches. User profiling as a service would make it possible for the user to make conscious decisions, who is allowed to profile their data, and at the same time enable data monetization - as transaction fees can be imposed when data is collected from the searched profiles. Users can also limit the information that is showing up in searches and demand the usage of pseudonymized data. User discovery service makes the user profiling an easily available tool for companies to be used for example in targeted marketing. Research institutes would be able to use the service to do population based research. For the human user the benefit is in transparency of the profiling and the opportunity to expose data for wider use, even sell it.

Local Applications

Local application refers to programs that do not transmit user data onto another Web server i.e. don't need a data connection for actual data transfer. Instead the application or data analyzing code is downloaded and run locally beside the data. Local applications may work in the user's personal data store (own computer or dedicated cloud service) or on a user controlled server where the data is originally stored (Data Source).

MyData Standards and Specifications

Ideas for standards and specifications help in developing interoperability between services and applications in the MyData architecture.

Data Resource Descriptions, Taxonomies and Data Resource Discovery

Data Sinks require data from suitable Data Sources to enable the service they provide. To facilitate the discovery of suitable Data Sources, a Data Resource Discovery functionality should be defined. This requires Data resource descriptions for data models, ontologies, taxonomies and semantics to be defined. A connected Data Source is the original provider of the necessary descriptive resources, which is managed through the MyData Account (TBD in more detail).

By organizing the data in a unified way, the overall usability of the system increases considerably. With common taxonomies to describe the Data Sources and Sinks, different services can be made more explicit for machine intelligence, which in turn facilitates service discovery, binding and interoperability. Data taxonomy is a hierarchical structure separating data into specific classes of data based on common characteristics. In MyData architecture, the following general aspects could be considered, for example:

- Temporal nature of data - is data collected as a continuous stream or periodically?
- Duration of storage of data - is data stored permanently or for a specified length of time?
- Stability of data - is data consistent and stable over time or is it dynamic by nature?
- Sensitivity of data - the level of sensitivity depends on whether access and handling of data is restricted or the data is available for anyone. Inappropriate handling of sensitive data could result in identity theft, financial loss or invasion of privacy, for example.
- Is data discrete or continuous – depends on whether data can only take particular values or take on any value within a finite or infinite interval.
- Rawness of data – does the Data Source contain e.g. numerical data collected directly from sensors or is it processed and analyzed data.

As stated earlier, the purpose of a data taxonomy is to establish a common understanding of technical usage guidelines of various data elements, which in turn enables ensuring correct interpretation of the data by its owners and end-users. To achieve this common understanding, certain characteristics and attributes of data elements have to be defined in specific metadata descriptions. The metadata definitions must reflect and be consistent with the taxonomy created. MyData Operators or other data brokers must make certain that this information is collected when Data Sources are registered via data APIs. Proper communication on the characteristics of available data facilitate the adoption and utilization of data in new services and applications. In technical level, the metadata descriptions may also be obtainable by computational systems, therefore the machine readable formatting of metadata should be available using e.g. JSON, XML or WSDL. In addition to physical data characterisation, semantic description must also be available. The minimum is a free-text description where the semantic interpretation of data is described so that it allows individuals to gain understanding what the data is about, how it is obtained and what elements it contains.

Service APIs description for Data Utilizers & Refineries

Data Resource Description can be used to describe the characteristics of Data Sources and Sinks in MyData Services. This will be translated into MyData Service API descriptions that are used to characterize data utilizers and refineries in the tasks of resource discovery and data exchange. A possible future use for API descriptions is to facilitate the exchange of data between two or more big data repositories. This exchange necessitates technical design that accounts for new big data system paradigms regarding extracting, transforming and loading of data into distributed storage, as well as processing and provisioning of data.

Consent Commons

Consent Commons refers to a long term goal of standardizing the contents of consents to a set of well known rights and use cases. Currently consents given to various data controllers are heterogeneous. However, the consents often contain similar kinds of elements that could be formatted among standard guidelines. When standardized, the consents can be made machine-readable and easy to compare, bundle, visualize and process automatically. Example can be taken from Creative Commons licensing framework which harmonized equally heterogeneous sphere of author rights to a common set of standard licences.

Registration of Actors to MyData Operator

In order for an Actor (Data Source or Data Sink) to be available for Account Owner to use, information about this Actor must be provided to MyData Operator. Although the Service Contract Template (SCT) and Consent Receipt structures presented in chapter 3 define some of the required information, the complete set of required information nor how this information gets recorded to the MyData Operator has not yet been defined. So a Service Registration Process that can require human interactions between the Actors, or dynamic and automated process where Actors (MyData Operator, Data Sink, Data Source) exchange information needed to establish a Service Connection should be defined. Protocols similar to proposals *'OAuth 2.0 Dynamic Client Registration Protocol'* and *'OpenID Connect Dynamic Client Registration'* present a good example of such mechanisms.

Appendix 1: Glossary

Account Management Service (UI) / Tilinhallintapalvelu

User interface and related service for managing MyData Accounts, Service Connections and MyData Authorizations. This service may be offered by MyData Operator (organization taking the operator role), or the individual may run the the software herself (operator-less accounts or self-operator).

Account Owner (role) / Tilin omistaja

Account Owner is the person who originally created MyData account. Depending on the account type the owner may be strongly authenticated or even anonymous. In case of individuals, the Account Owner usually is the same as the data subject.

Actor ID / Toimijan ID

A string that uniquely identifies the actor.

Consent (legal base) / Suostumus

Consent is the permission to process personal data and it is always given by the Data Subject to the Data Controller. To be legally valid, the Data Subject needs to give an unambiguous consent to the processing of their data to specific purposes, and the consent has to be free, informed and explicit. Consents can be part of a larger (not just MyData-context) legal contract between the individual and services she uses.

Consent Receipt (record) / Suostumuskuitti

A pair of Consent Receipts (one for the Source and one for the Sink) is created as a result of a MyData authorization. For the Source it defines, what data can be provisioned to the specified Sink, and for the Sink it defines, how the data can be accessed and processed. A Consent Receipt is a manifestation of legally valid Consent and makes it technically feasible to change or withdraw the consent dynamically.

Data Connection / Datayhteys

Authorized transfer of data from Data Source to Data Sink.

Data Controller (legal role) / Rekisterinpitäjä

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.

Data Processor (legal role) / Datan käsittelijä

A natural or legal person, public authority, agency or any other body which processes personal data on behalf of the data controller. A role that is not defined in U.S. privacy legislation so far.

Data Subject (legal role) / Datasubjekti tai rekisteröity

An identified or identifiable natural person whose personal data is processed. The data subject (individual) should have rights and practical means to control creation, flow and usage of his personal data. The data subject gives and manages Consents related to his data and Service Connections.

Data Sink (role) / Sinkki

Entity that can capture data from one or more Data Sources and allows management of data usage through a MyData compliant API and uses this data to produce new applications, services and information.

Data Source (role) / Datalähde

Entity that can provision data about the Account Owner to one or more Data Sinks and allows management of data provision through MyData compliant API.

Data Usage Log (records) / Datan käytön logi

Occurrence of Data Connections between Data Source and Data Sink are recorded in a Data Usage Log stored at MyData Account.

MyData / MyData

Subset of personal data. All MyData is personal data, but such personal data where the data subject has no practical means to get the data for herself and control how it is used can not be called MyData.

MyData Account / MyData Tili

MyData account hosts all Service Contracts and Consent Receipts that define access means and rights for Data Sinks and Data Sources.

MyData Authorization (interaction) / MyData luvitus

An authorization proves there exists Account Owner's permission for active data transfer from a specific Source to a specific Sink. When authorization takes place a pair of Consent Receipts (one each for the Source and the Sink) are created.

MyData Compliant API / MyData yhteensopiva API

Data sources and Sinks need to provide API which gives unique endpoint address recognizable for other MyData compliant services. Functionally the API needs to support 1.) access management through Service Connections and MyData Authorizations 2.) import and/or export of data through authorized Data Connections.

MyData Operator (role) / MyData operaattori

MyData (consent) operator provides MyData Accounts and the related Account Management Service.

Personal Data (definition) / Henkilötieto

All kinds of data related to the person or resulting from the person's activities. The term covers more than just the most personal information, such as name and address. It is also not strictly limited to the legal definition of personally identifiable information.

Service Connection (interaction) / Palvelun linkitys (MyData tiliin)

Action executed by an Account Owner to link a service (Data Source or Data Sink) to her MyData Account. As the result the Service Connection status and parameters are documented within a digital machine-readable record, called a Service Contract.

Service Contract (record) / Palvelusopimus

Service Contract is the formal outcome of a successful Service Connection. It documents in machine readable form the terms and scope of the agreement between the Individual (Account Owner) and a single Service (Data Source or Data Sink). Service Contracts are stored at the related MyData Account.