MIM 4 – PERSONAL DATA MANAGEMENT SPECIFICATIONS

Personal Data Management

Minimum Interoperability Mechanism (MIM) 4

Introduction and Specifications

Final draft

May 2021

This MIM4 specification was created in the **City as MyData Operator** project coordinated by the City of Helsinki and with participation from City of Espoo, City of Turku and City of Oulu. The project is funded by the Ministry of Finance in Finland.

This version was actively contributed by Sami Jokela (1001 Lakes Oy), Marko Turpeinen (1001 Lakes Oy), Olli Pitkänen (1001 Lakes Oy), Harri Honko (Vastuu Group), Ville Lavonius (Vastuu Group), Mika Huhtamäki (Vastuu Group), Jami Haavisto (Vastuu Group), Pirkko Laitinen (Vastuu Group), Matthias De Bièvre (Visions SAS), Mika Leivo (City of Helsinki) and Mikko Rusama (City of Helsinki).

Personal Data Management

Minimum Interoperability Mechanism (MIM) 4

Introduction and Specifications

Suggested citation

Please refer to this document using the following notice:

Personal Data Management, Minimum Interoperability Mechanism (MIM) 4, Introduction and Specifications by 1001 Lakes Oy, Vastuu Group, Visions SAS, and City of Helsinki, May 2021.

[1] Some figures in this paper utilize and are derivative works from "MyData Graphics" icons by MyData Global, licensed under CC BY 4.0.

MIM4 Specifications – Personal Data Management

Summary

Citizen data is a key element in providing human-centric public services. Open and Agile Smart Cities (OASC) explores the mechanisms for personal data protection, transparency, and trust when sharing personal data between cities. These capabilities are a key part of the digital transformation journey of cities worldwide. The journey is supported by common and open standards and open technical specifications like the Minimal Interoperability Mechanisms (MIMs) that more than 150 cities in over 30 countries have adopted. The objective is to have cities and communities replicate and scale data sharing solutions globally.

This version of Minimum Interoperability Mechanism for Personal Data Management, MIM4, approaches personal data management and connectivity through two pillars. The first pillar describes an open-source access gateway – a connector that enables multiple cities to utilise one Data Source without needing to build their own Data Sources or connectors. The second pillar is a legal framework that governs the use of connectors for data access.

Table of Contents

PART 1 – MIM4 Operating Context

1 Personal Data in Smart City Context

1.1 Introduction to Personal Data Management in City Context

Through digitalisation, cities strive to provide better and easier-to-use services to their residents. Cities are also major producers and users of data. The urban data spectrum covers the built environment, transport, education, health, hobbies and the personal preferences of urban dwellers. Thus, data covered by urban activities is a very valuable resource that can be used to improve our daily lives.

As public actors, cities should promote a more people-centered digitalisation, where citizens have better means of managing the use of their personal data in urban services without compromising privacy. From a people-centered perspective, the sharing of personal information is based on trust and balanced and fair relationships between individuals and organizations. Citizens must be able to easily give and withdraw consent to the processing of their personal data for a specific limited purpose, when legislation does not oblige, for example, a public authority to process such data. In any case, people need to be informed about the processing of personal data and they need to be able to understand how their data is used. In order to allow the exchange of information, it is necessary to agree on common rules and interoperability mechanisms for the sharing of information between the different parties.

The aim of this specification is to define common approaches to the movement of urban data between different cities in a manner that is fair and transparent to the citizen, thus significantly increasing the potential for the utilization of urban data. In many use cases, data must be able to be exchanged between cities, e.g., shared service production organized as an urban joint venture or organization, a city-managed benefit, or a city-managed training and competency information.

The benefits of common rules and agreed interoperability mechanism can be seen from the viewpoints of the city, the citizen or the service provider:

- **Benefits for citizens:** Digital transactions involving personal data are often complex - information is not easily transferable between services and has to be entered many times. Inconsistent approaches to requesting and using personal information undermine understanding and trust. Sharing data between cities supports life situations where citizen data needs to be moved between cities, for example, moving from one city to another. Better and smoother proactive and personalized online services improve the quality, efficiency, and availability of online services, and trust in the services and the city will increase.
- **Benefits for cities:** The potential of personal data is not currently fully exploited because of lack of interoperability with policies, practices and standards for the flow of information between services / cities. Interoperability mechanisms reduce bureaucracy as personal information becomes more readily usable in services. Interoperability mechanisms provide an integrated approach to speed up the procurement and implementation of services and increase confidence and satisfaction in online transactions.
- **Benefits for service providers:** Urban personal data is often difficult to use in development of services created by the private sector, although this could be desirable. Interoperability mechanisms result in businesses having easier, trusted and secure access to citizen data, data being available in a form that can be used flexibly, and assurance as to what information may be

authorized to be used. Common specifications to be created and implemented will create a more efficient market in which companies can develop urban services using personal data.

1.2 Legal Framework

From the legal point of view, personal data management is governed by legislation on one hand and by contractual arrangements on the other. Below, the most important legislation in relation to personal data management is described briefly, and an overall picture about contractual arrangements is given.

On the legislative level, the data protection law is very important background for this specification. An essential part of it is the EU's General Data Protection Regulation (GDPR, 2016/679). As a general law, it applies to almost all the processing of data involving personal data. However, the regulation of personal data protection is quite fragmented. In addition to the General Data Protection Regulation, the processing of personal data is regulated by numerous other laws both on the EU level, such as ePrivacy Directive (2002/58/EC), which will be replaced by ePrivacy Regulation in the future, and on the national level local adaptation laws, like the Data Protection Act in Finland, as well as specific laws, like those about data protection in employment relationships, electronic communications services, education, healthcare, and so on. It is not possible to take a position on all of them in this specification. Where possible, the main data protection provisions for the services have been taken into account, but the users of the specification should not trust that it covers all the relevant legislation, but they always need to check that the applicable laws are obeyed.

According to the GDPR, 'personal data' means any information relating to an identified or identifiable natural person ('data subject').[2] It is important to notice that this includes not only information that identifies somebody or is directly linked to someone, but the definition also includes information that is related to an indirectly identifiable natural person. It means that information is personal data although it is not directly linked to a human being, but it is possible to link it to someone, if other information is combined thereinto. According to the European Court of Justice that would not be the case if the identification of the person was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and manpower, so that the risk of identification appears in reality to be insignificant.[3] Despite these limits, the definition of personal data is anyhow very broad. Therefore, it is often the safest to assume that any data may include personal data, unless it is absolutely sure that no data can be linked to any individuals.[4]

Examples of personal data:[5]

- E-mail address, such as firstname.lastname@company.com

[2] GDPR Art 4(1).

[3] ECJ Case C-582/14.

[4] Wrigley, J. S. P., Alen-Savikko, A. K., & Pitkänen, O. (2019). Finding the 'Personal' in the Industrial Internet: Why Data Protection Law still Matters. In R. Ballardini, O. Pitkänen, & P. Kuoppamäki (Eds.), Regulating Industrial Internet Through IPR, Data Protection and Competition Law (pp. 235-252). Kluwer Law International.

[5] https://tietosuoja.fi/en/what-is-personal-data

- Telephone number
- Identity card number
- Car registration number
- Positioning data (e.g., from a mobile phone)
- IP address
- Patient records
- A pet's veterinary records
- Data on the hereditary diseases of the person's great-great-grandparents

However, it is important to remember that any data can be personal data if it is related to a certain human being. For example, "185 cm" is usually not personal data, but if it is combined with other information that this measure indicates how tall a certain person is, it becomes personal data. A street address as such is hardly personal data, but if we are able to find out, who lives there, it turns out to be personal data.

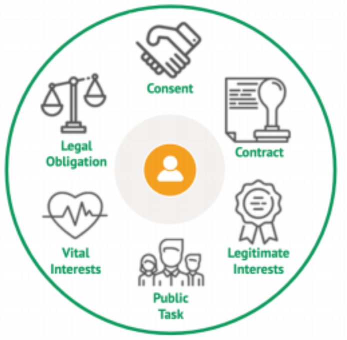


Figure 1: Lawful bases for processing data based on GDPR[6]

The GDPR Art 5 defines several important, yet rather vague principles that must be followed when processing personal data. They include that personal data shall be:

- processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes ('purpose limitation');
- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
- accurate and kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate are erased or rectified without delay ('accuracy');
- stored for no longer than is necessary for the purposes for which the personal data are processed ('storage limitation');
- processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction,

[6] MyDataShare White paper v1.0.1, Vastuu Group Oy. Available at https://kampanja.vastuugroup.fi/hubfs/MyData/MyDataShare-Whitepaper-v1.0.1.pdf

or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

- The controller shall be responsible for, and be able to demonstrate compliance with, the previous principles ('accountability').

So, in order to process personal data, there must be a lawful basis, which according to the Article 6 of the GDPR is one of the following:

- an informed consent of the individual,
- a contract with the individual,
- a legal obligation,
- a vital interest of the individual,
- a public interest,
- a legitimate interest that is not overridden by the interests or fundamental rights and freedoms of the data subject.

It is a good idea to apply as few bases in one case as possible. In relation to personal data management, consent is often thought to be the most important lawful basis, but it is important to realize that the other bases may also need to be considered. Especially, in a city-context, a legal obligation or a contract can often replace consent as the lawful basis. On the other hand, vital interest can be highly relevant especially in healthcare and emergency services, while public interest can be the most applicable basis for processing information e.g. about decision makers, and especially commercial companies may justify the processing of personal data in many cases by referring to a legitimate interest.

Data cannot be transferred from the data provider to the data consumer without a valid permission. A MyData Operator explained in the following Chapters maintains information on the lawful basis for the data processing.

Besides the data protection law, there are numerous other statutes to be considered in relation to personal data management. In the city-context, they include especially national laws governing e.g. city administration, information management in public administration, and the openness of government activities.

Furthermore, in addition to the legislation, contractual arrangements are essential to govern personal data management from the legal point of view. The most important agreements within this specification's scope belong to two groups:

- Contracts with the individuals, whose personal data is processed. Those agreements may include provisions that make the processing of personal data necessary. Therefore, they can establish a lawful basis for the processing of personal data in accordance with the GDPR Art 6(1)(b). On the other hand, solely the existence of an agreement with an individual does not necessarily in itself guarantee that there is a lawful basis for processing personal data and thus it is still important to consider case by case if another lawful basis is required.
- Contracts between any actors, like operators, data providers, service providers, and data users, participating somehow in data sharing and processing. These contracts may be data processing agreements between controllers and processors, service agreements on infrastructure services, project agreements or any other agreements that include provisions about processing personal data.

In general, these agreements are called *data sharing agreements*. They can be simple bilateral contracts, but in a multi-party network, it is often preferrable to turn them into rulebooks, like Sitra's

Rulebook for a fair data economy.[7] An agreement may also be machine-readable to enable more or less automated contractual relationship management as described later in this specification.

1.3 MyData as a driving principle for personal data management

Central to MyData[8] thinking is individual's ability to control the use of information about him or herself. The MyData approach aims at strengthening digital human rights while opening new opportunities for businesses to develop innovative new services based on personal data and mutual trust.

Key principles of MyData are summarized as follows:

- **From formal to actionable rights.** We need to move from friction to action, by making exercising rights simple and easy.
- **From data protection to empowerment.** We need to move from fear to confidence by understanding how sharing data can be good for the individual.
- **From closed to open ecosystems.** We need to move from monopolies to freedom of choice among good alternatives through openness.

Key actors in the MyData ecosystem according to MyData definitions are Data Source, Data Using Service, Operator, Person, as depicted in the following figure.

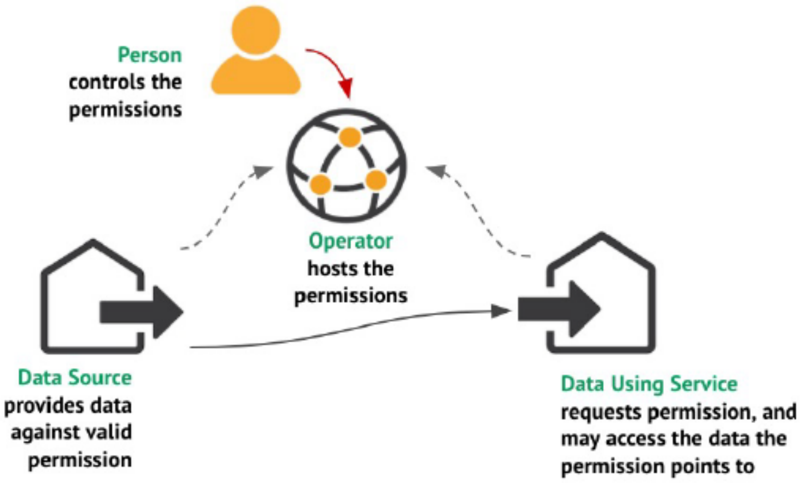


Figure 2: Key parties in MyData-based personal data management [9]

- **Person:** The role of data subject as represented digitally in the ecosystem. Persons manage the use of personal data about themselves, for their own purposes, and maintain relationships with

[7] https://www.sitra.fi/en/publications/rulebook-for-a-fair-data-economy/

[8] MyData Global (http://mydata.org)

[9] MyDataShare White paper v1.0.1, Vastuu Group Oy, icons used under courtesy of MyData Global. Available at https://kampanja.vastuugroup.fi/hubfs/MyData/MyDataShare-Whitepaper-v1.0.1.pdf

other persons, services, or organisations. The words **Individual** and **Citizen** are also used in this document as a synonym for a Person.

- **Operator:** The role responsible for operating infrastructure and providing tools for the Person in a human-centric system of personal data exchange. Operators enable people securely to access, manage, and use personal data about themselves as well as to control the flow of personal data within and between Data Sources and Data Using Services. In this document, we use the term MyData Operator to reflect an entity acting in this role.
- **Data Source:** The role responsible for collecting, storing, and controlling personal data which persons, operators, and Data Using Services may wish to access and use. In this document, Data Source is synonymous with the term **Data Provider.**
- **Data Using Service:** The role responsible for processing personal data from one or more Data Sources to deliver a service. In this document, Data Using Service is synonymous with the term **Data Consumer.** When highlighting the aspect that the service utilizes personal data using MyData Operator as the Operator, we use the term **MyData service**.

MyData Operators are responsible for the infrastructure and tools that enable secure access and control of the flow of personal data within and between Data Sources and Data Using Services. These operators should be able to share personal data easily, both in and outside of their organisations - on both technical and legal levels - without an extensive amount of legal work and point-to-point integrations.

According to MyData Operator White Paper[10], MyData Operator has two main directions in which they operate:

- **For individuals:** Operators provide transparency, understandability, and convenience to individuals when they share data or receive services using data about them. Operators provide an aggregated view to an individuals' personal data, allow them to control who can use the data and for which purpose, and transparently expose past data use and sharing. Other benefits include intuitive user interfaces, enhanced security, and the tools for managing relationships with different services that process personal data
- **For organisations:** Operators provide easy, legally compliant connectivity to an ecosystem of Data Sources and Data Using Services as well as a relevant base of potential users. Operators facilitate access to high quality, up-to-date data in real time, offer tools and mechanisms for legal compliance such as logging and audit trails of permissions, and offer outsourced tools for complying with data portability requirements.

MyData principles for operators include:

- **Human-centric control of personal data:** This principle requires that any personal data transaction by an operator always involves the individual. There are legitimate purposes for using personal data that do not require specific consent from the concerned individual, but also for these transactions there should be transparency and availability to check how personal data

[10] Adapted from https://mydata.org/wp-content/uploads/sites/5/2020/04/Understanding-Mydata-Operators-pages.pdf

has been utilized. It also requires that the actions required of and performed by the person, such as giving permission, are very easy for individuals to understand.

- **Individual as the point of integration:** Operators deliver the integration of services and data to the individual and, therefore, have a responsibility towards the individual (a duty of care).
- **Individual empowerment:** This principle requires operators to support a shift from an individual merely giving permissions when asked, to them having a wide range of real choices, the initiative regarding data about them, and the ability to negotiate terms.
- **Portability – access & re-use:** This principle allows individuals to go beyond control of their data to create their own uses for personal data. Operators must support individuals to re-use personal data about them.
- **Transparency & accountability:** Adopting these principles, operators must be prepared to deal with intended as well as unintended consequences of personal data use in a manner that creates trust and mitigates potential risks. Without transparency, personal data sharing practices cannot be inspected or contested.
- **Interoperability:** Interoperability requires that individuals are able to move between operators and to transfer data within the ecosystem without the need for transformation or interpretation. Operators must work together, and with other actors, to achieve this.

1.3.1 Interoperability mechanisms according to MyData

The interoperability requirement for a MyData Operator is to describe the systems for personal data management with respect to the MyData Operator reference model.

Interoperability has the following focus areas:

- **Transparency and usability:** Turning formal rights into actionable rights for people. This means using control vocabularies and semantics for transparency and common elements of user experience such as recognisable icons and labels.
- **Standardising interfaces for personal data:** Enabling ecosystems to scale fast and for data portability to become seamless.
- **Enhancing roaming possibilities:** Enabling the routing of data transactions via multiple operators so that there is no need for all people and all services to link to a single operator.
- **Enabling substitutability:** Supporting easy switching of operator services and, ultimately, fungibility of base functionalities which are entirely interchangeable with indistinguishable inputs and outcomes.

Interoperability can be discussed at different dimensions, such as:

- **Technical level:** Definitions of connectivity, syntactics, and protocols for data exchange (e.g., APIs) and data storage that underpin basic integration. The first objective here is to enable the easy connection of new Data Sources and Data Using Services to an operator and their mutual interoperability, where operators can work with each other technically.
- **Semantic level:** Harmonised information with shared data models and mutually agreed content. The pragmatic approach here is to identify the categories of data where common data models are most essential for MyData. These could be semantic models for data control and governance

(e.g. data transaction records, consent records purpose categories) or widely used attribute data types and domain specific data models.

- **Organisational level:** interoperability in more mature ecosystems goes beyond the technical and semantic levels, encompassing shared objectives and policies between organisations. These objectives and policies will cover issues such as responsibilities, liabilities, business models, and governance structures.

1.3.2 MyData Operator business models

As discussed earlier, benefits from the ecosystem participation should be higher than the costs. This includes the role of the MyData Operator acting as intermediary in personal data transactions. Some applicable sources and business models for MyData Operators include:

- **Person:** one-time onboarding fees, recurring account fees, or pay-as-you-go fees.
- **Other operators:** roaming fees, or a share of transaction and connection fees.
- **Data Source:** one-time onboarding fees, recurring account fees, or sales commission.
- **Data Using Service:** one-time onboarding fees, recurring account fees, transaction fees, or connection fees.

In the city context, it is likely that the costs of the operations could be covered by the city, as part of the municipalities' mission to provide high quality, trustworthy and secure digital services.

1.4 Framework for Personal Data Management for Cities

The operating model for Personal Data Management in an urban environment can be seen as a multi-level framework as depicted in the following figure.

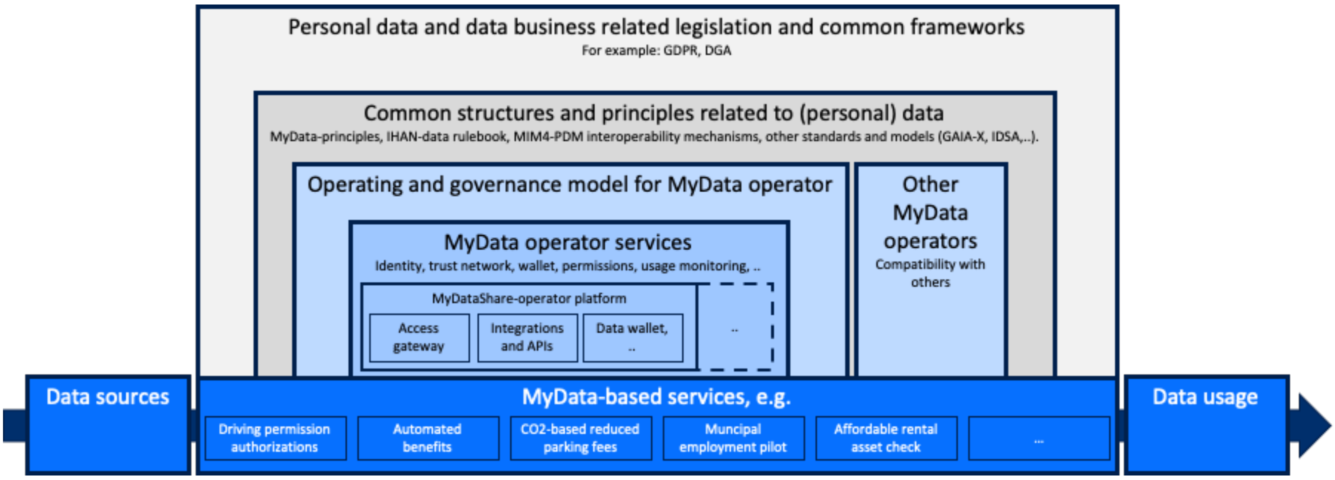

Personal data and data business related legislation and common frameworks
For example: GDPR, DGA
Common structures and principles related to (personal) data
MyData-principles, IHAN-data rulebook, MIM4-PDM interoperability mechanisms, other standards and models (GAIA-X, IDSA,..).
Operating and governance model for MyData operator
MyData operator services
Identity, trust network, wallet, permissions, usage monitoring, ..
MyDataShare-operator platform
Access gateway
Integrations and APIs
Data wallet, ..
..
Other MyData operators
Compatibility with others
Data sources
MyData-based services, e.g.
Driving permission authorizations
Automated benefits
CO2-based reduced parking fees
Muncipal employment pilot
Affordable rental asset check
...
Data usage

Figure 3: Personal Data Management in smart city context

At the top level is general legislation, such as data protection legislation, and supported by general personal data management models, such as the MyData principles. Together, they define both the operation and the management model of the data intermediary, such as a MyData Operator, as well as the network of these data intermediaries. The focus of this network is on the portability of personal data and access rights between cities. At the core are the MyData Operator services provided to various actors implementing the services, on the basis of which a growing number of use cases related to the utilization and transmission of personal data are implemented. The examples mentioned in the above figure are possible MyData Operator services in an urban context.

It is important to stress that MyData Operator services provide mechanisms for personal data usage control covering several possible lawful bases for processing data (citizen's consent, employment contract, legitimate interest, etc.). Also, MyData Operator does not handle the actual transfer of data from the Data Source to Data Using Service, i.e. the data does not "flow through" the MyData Operator.

To summarize, the key functions provided by the MyData Operator in the city context, are to (1) check whether the Data Using Service is permitted to use the Data Source, and (2) provide transparency by making personal data usage visible to the Person whose data has been used.

2 Minimum Interoperability Mechanisms (MIMs)

Open & Agile Smart Cities (OASC) is a non-profit, international smart city network aimed at creating a data and services market for smart cities. The mission of OASC is to unite cities and communities around the world to build a global market for solutions, services, and data based on the needs of cities and communities. To achieve this mission, OASC champions the Minimal Interoperability Mechanisms (MIMs), a set of practical capabilities based on open technical specifications that allow cities and communities to replicate and scale solutions globally.

MIMs are common tools for data, systems and to achieve interoperability of services between cities and service providers. They provide the technical foundation for procurement and deployment of urban data platforms and end-to-end solutions in cities & communities worldwide.

Three MIM definitions have been published and introduced for pilots:

- MIM1: Context Information Management
- MIM2: Common Data Models
- MIM3: Marketplace Enablers (Ecosystem Transaction Management)

Two additional MIMs have been proposed and accepted as work items:

- MIM4: Personal Data Management
- MIM5: Fair Artificial Intelligence

The MIM4 is thus part of a larger, OASC-coordinated set of MIMs. Each MIM specification is led by a specific city, and behind each MIM definition is a set of experts who are responsible for development. City of Helsinki has taken the lead in developing the MIM4 for Personal Data Management.

3 MIM4 – Personal Data Management

MIM4 creates practices for the movement of personal information between different services and cities with the consent of the citizen, thus increasing the potential for the utilization of urban information significantly. It is expected to provide a clear and easy usable means for citizens to take control of their personal data and to control which services can access and use their data sets or attributes.

The main **objectives** of the MIM4 are:

1. Support service ecosystems globally to create unified practices for personal data management and usage in an urban context by creating a personal data interoperability mechanism.
2. Support the service ecosystems in cities adhering to MIM4 specification to boost the creation of a common international market for urban services that utilize personal data.
3. Increase the ease of use of urban digital services for citizens by enabling a smooth user experience in service chains across cities
4. Increase trust between citizens and service providers by reducing data protection risks and worries about misuse of personal data.

Core **outputs** of MIM4 specification effort are:

1. Description of the context and scope for the MIM4 - Personal Data Management for smart cities and its relation to other minimum interoperability mechanisms, MIMs.
2. Description of MIM4 for implementing personal data management in city context, driven by identified use cases and focusing on the identified key aspects of the interoperability.
3. Provide a concrete, open-sourced implementation of the personal data management referred in the specifications.

MIM4 Personal Data Management needs to have an open API in line with MIM1 to broker data and standard data models MIM2. The MIM4 needs to be fully compatible with the GDPR and needs to enable users to handle consent, allow and revoke access, and have full transparency on their personal data. Thus, MIM4 supports the right of citizen to have insight what personal data is available, stored, shared, etc. by the providers of the applications and/or services in use. Depending on the lawful basis of data processing, the citizen may also have the right to change and/or delete part or all personal data available, stored, shared, etc. by the provider of the applications and/or services in use. MIM4 also allows users to choose their MyData Operator and have a freedom of choice in moving their data and permission management to the operator of their choosing.

Based on the MyData principles and interoperability requirements, we can define four levels of personal data management as follows[11]:

- **Layer 1: Identity and trust networks.** A citizen profile or account to which we can attach personal digital identities, data preferences and data sharing agreements.
- **Layer 2: Management of data sharing agreements.** A way to define data sharing agreements that define data, purpose, and lawful base for processing.
- **Layer 3: Personal data platforms.** The standard should only define the APIs (e.g., for Layers 1 and 2) and each MyData Operator could then integrate their own backends and/or purchase compatible operator capability as a service as they see best.
- **Layer 4: Personal data models.** A shared personal data model OR a way to map distinct models to each other.

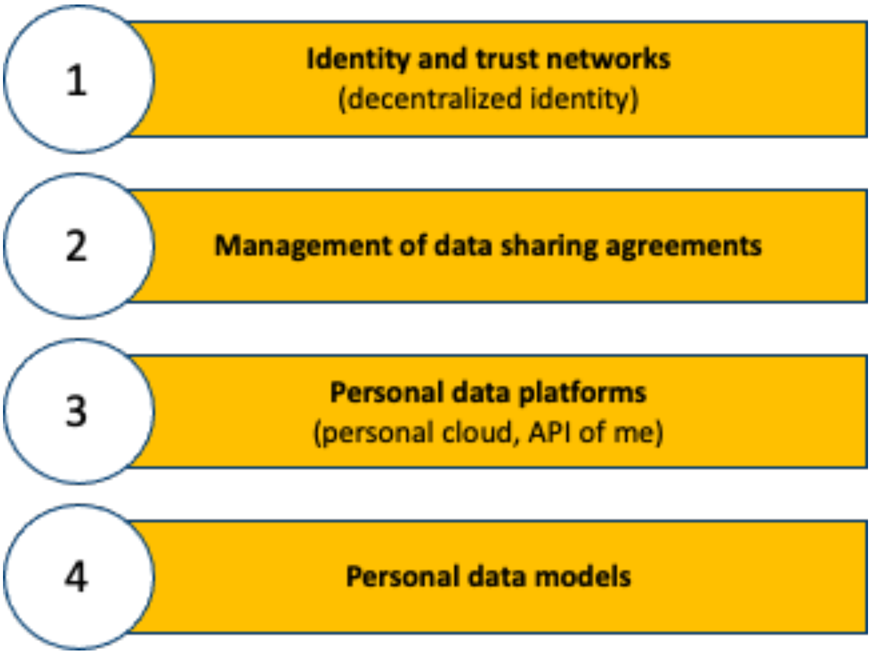

1
Identity and trust networks
(decentralized identity)
2
Management of data sharing agreements
3
Personal data platforms
(personal cloud, API of me)
4
Personal data models

Figure 4: MyData-based requirement layers for personal data management

To allow citizen to truly "roam" between cities and their MyData Operators, a few other capabilities are needed:

- **Import/export:** To transfer all permission-related data from one operator to another
- **Delete:** Citizen should be able to delete all their consent-based data or at least prevent the use of it.

[11] https://living-in.eu/group/7/commitments/mims-plus-technical-specifications-v3

- **Archive:** Legislation may require that some data is archived and cannot be deleted

This version of the MIM4 specification has its focus on data connectivity related interoperability – providing first a technical mechanism to open the Data Using Services of different cities a technically straightforward access to Data Sources residing in different MyData Operator's domain.

The following figure describes the scope of the MIM4 specification, where the focus in on providing a standardised way to connect a Data Source to a MyData Operator and further to a Data Using Service.

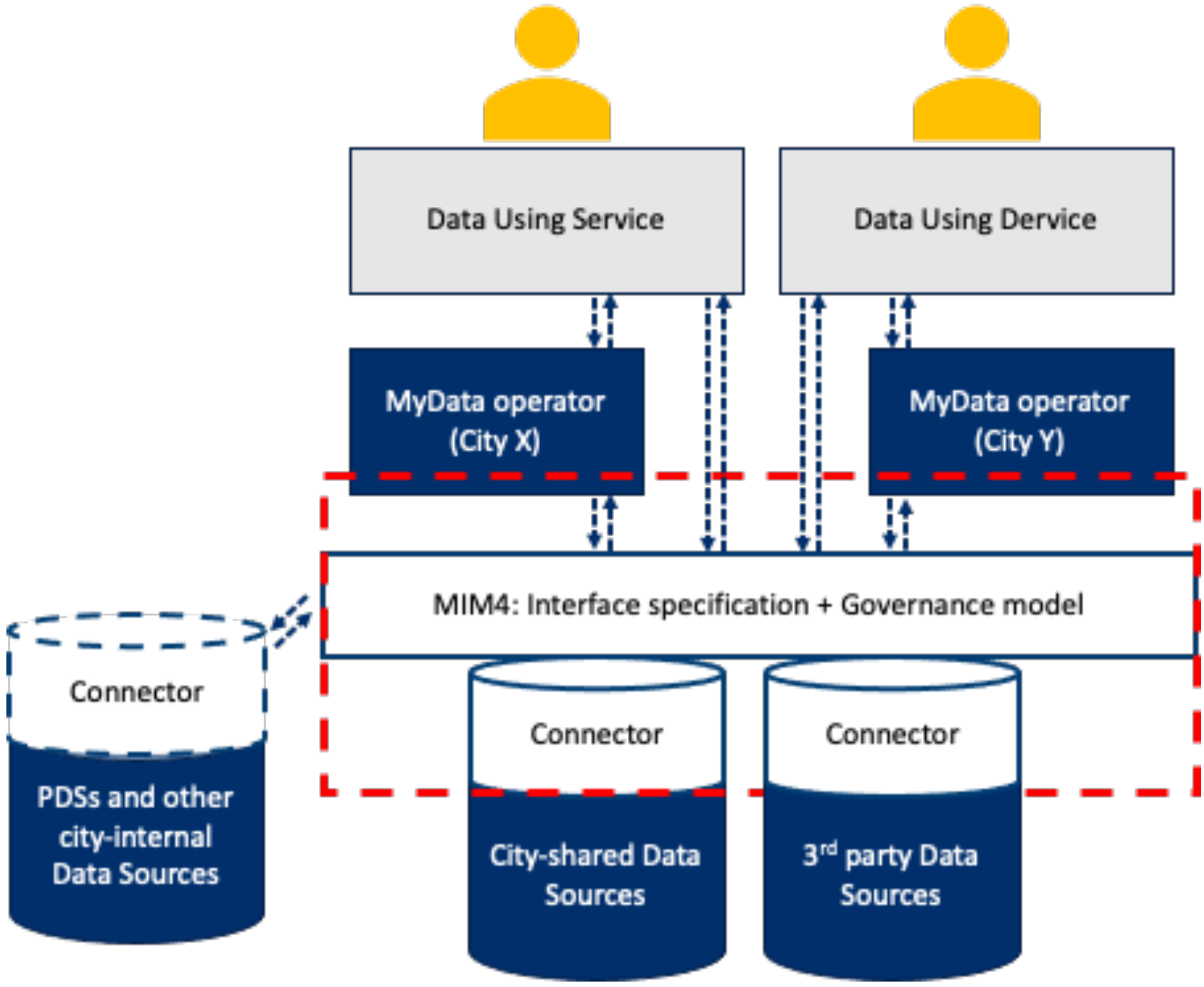


Figure 5: Scope of MIM4 focusing on connecting Data Source to a MyData Operator

Key **benefits** of MIM4 specification are:

1. Standardize the connectivity layer via connectors for citizen data sharing so that cities and other parties can develop compatible personal data management solutions available across multiple MyData Operators.
2. Standardize the legal layer to set, check and enforce data sharing agreements between parties and to define access and usage control based on those agreements.
3. Minimize the required development effort for implementing MyData-compatible Data Using Services by using the MIM specifications and available reference implementation.
4. Create an interoperable solution for Data Sources to make personal data available under strict control and permission management.
5. Provide means for realizing multi-operator trust framework between MyData Operators.
6. Reduce risk of operator lock-in for the Data Using Services and Data Sources.
7. Simplify Data Consumer access to Data Sources through a legal trust framework and shared connectivity layer.
8. Improve interoperability between multiple MyData Operators.

3.1 Other specifications and frameworks related to MIM4

In addition to the MyData Operator model described in the previous chapter, there are other reference models that are relevant when considering the MIM4 specifications, although they are not discussed here more in detail. Potential other frameworks and reference include:

- IDS Reference Architecture Model [12]
- IHAN as testbed for Fair Data Economy [13]
- SOLID [14]
- IRMA [15]
- Trust Over IP (ToIP) [16]

4 Key capabilities and processes

This document focuses on personal data management related to city-related services. In these cases a city acts as a Data Using Service provider, a MyData Operator manages person's data and/or its permissions and one or more Data Sources act as a source of the personal data. Human-centric use of personal data means that the individual is actively involved in the process either through providing

[12] IDS Reference Architecture Model: https://internationaldataspaces.org/use/reference-architecture/

[13] IHAN testbed: https://www.sitra.fi/en/projects/testbed-for-fair-data-economy-ihanfi/

[14] SOLID: https://solidproject.org/

[15] IRMA: https://privacybydesign.foundation/irma-en/

[16] Trust Over IP: https://trustoverip.org/

access to individual's personal data or gaining visibility to personal data processing related to the individual.

The following figure describes the key actors related to the personal data management and their key activities using personal data management in service execution.

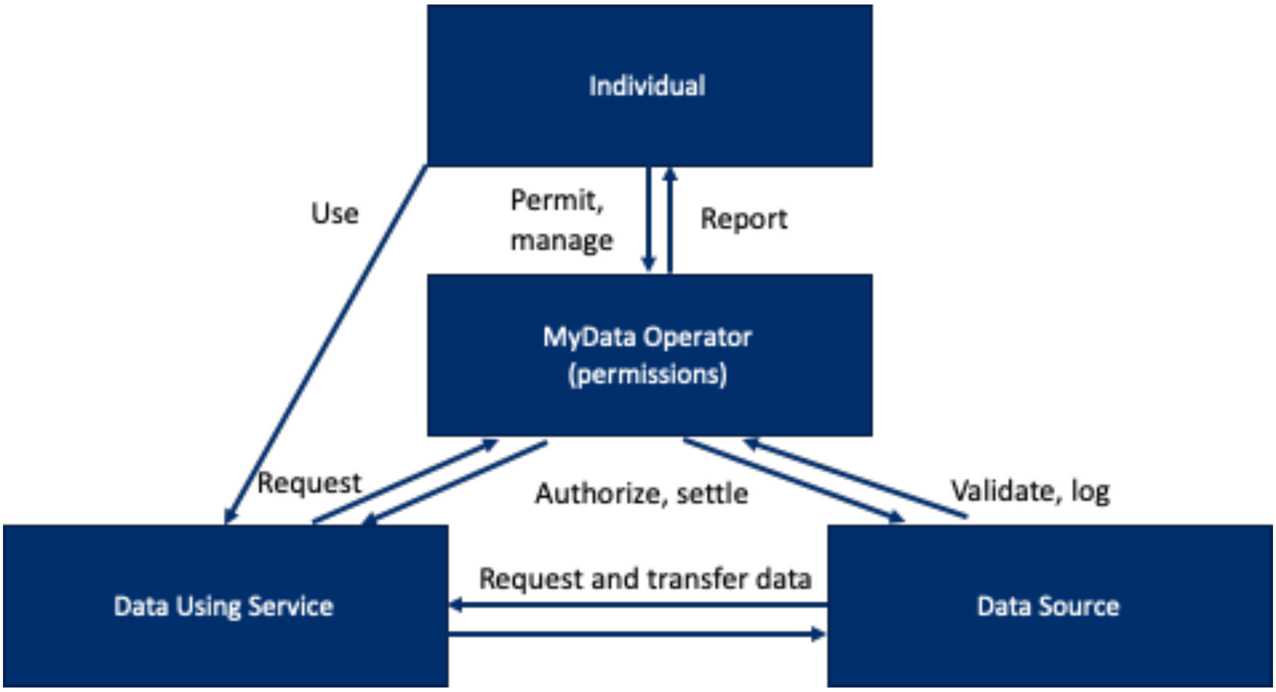


Figure 6: Business roles and interactions related to personal data management

4.1 MyData-based capabilities

MyData Operator reference white paper describes a number of key capabilities that a MyData -based operator should consider as part of the personal data management. These capabilities are illustrated in the figure below and include:

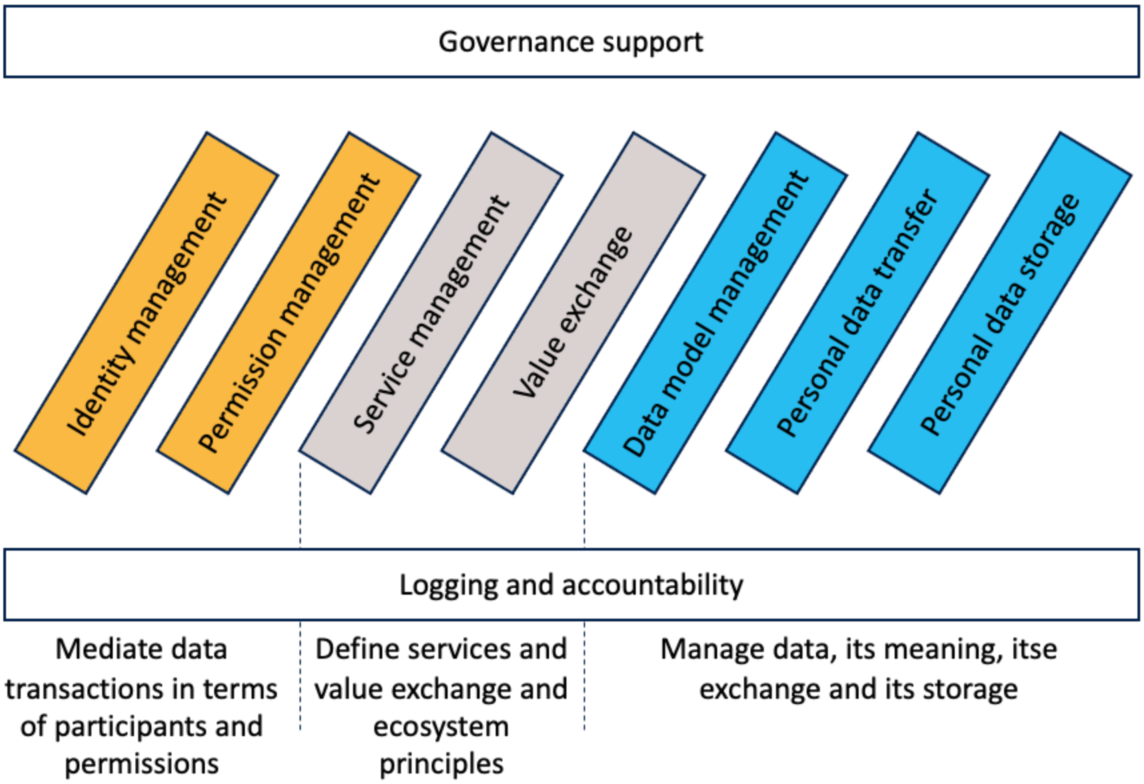

Governance support
Identity management
Permission management
Service management
Value exchange
Data model management
Personal data transfer
Personal data storage
Logging and accountability
Mediate data transactions in terms of participants and permissions
Define services and value exchange and ecosystem principles
Manage data, its meaning, itse exchange and its storage

Figure 7: Main functions for a MyData Operator[17]

- **Identity management** handles authentication and authorisation of individuals and organisations in different, linked identity domains and links identities to permissions.
    - Individuals can have different identities, or profiles, with different Data Sources and Data Using Services.
- **Permission management** enables people to manage and have an overview of data transactions and connections and to execute their legal rights. It includes maintaining records (notices, consents, permissions, mandates, legal bases, purposes, preferences etc.) on data exchange.
    - Focuses on human-centric control of personal data.
    - Includes also legal compliance for Data Sources and Data Using Services participating in the data transaction.
- **Service management** uses connection and relationship management tools to link operators, Data Sources, and Data Using Services. Data can be available from different sources and can be used by multiple Data Using Services.
    - Operator manages individual's permissions and links dynamically together Data Sources, Data Using Services, and individuals' data

[17] Adapted from Understanding MyData Operators. Available at https://mydata.org/wp-content/uploads/sites/5/2020/04/Understanding-Mydata-Operators-pages.pdf

  - Comprises of both access control and technical connection management. Delivery of these functions is agreed case by case with the operator having varying degrees of responsibility in the activities.
  - A significant decision for multi-operator environment is whether to use a shared registry or if each operator manages services separately.
- **Value exchange** facilitates accounting and capturing value (monetary or other forms of credits or reputation) created in the exchange of data.
  - All actors in the ecosystem need to have more benefits than costs.
  - Practical level functionality is to track transactions for value transfer, such as payments or other forms of rewards, in the form of logs and reporting. This might also include paying individuals for their data, but rationale or argumentation for this is not in the scope of the MyData Operator reference model.
- **Data model management** is about managing the semantics (meaning) of data, including conversion from one data model to another.
  - Focuses on translating semantics from one data model to another. Might also include interpreting standard data models to individuals. Some operators provide data harmonization while other are focusing more on pure data transfer without data interpretation.
  - Includes data model standardization supporting interoperability, such as log data syntax or permission models.
- **Personal data transfer** implements the interfaces (e.g. APIs) to enable data exchange between the ecosystem participants in a standardised and secure manner.
  - Key function related to data portability, access and re-use of personal data.
  - Allows data exchange between Data Sources, Data Using Services and operators in a standardized and secure fashion adhering to defined permissions.
  - Actual implementation can be with or without the operator as a middleman.
  - Includes also master data type of questions; how to avoid unnecessary duplication and easy updates of data.
- **Personal data storage** allows data to be integrated from multiple sources (including data created by a person) in personal data storage (PDS) under the individuals' control.
  - Replicates personal data from Data Sources and allows manipulation of further data by the individual so that the data is better under individual's control and Data Sources and Data Using Services are linked only via operator & individual.
  - "Personal master data" principle; single source of truth for personal data.
  - Personal data storage should be considered as a separate service and Data Source instead of being offered by the operator. In reality these two roles are often combined at operators.
- **Governance support** enables compliance with the underlying governance frameworks to establish trustworthy relationships between individuals and organisations.
  - Definition, monitoring, follow-through and administrating governance for the ecosystem.
  - Data policies and data sharing agreements for the eosystem.
- **Logging and accountability** entails keeping track of all information exchanges taking place and creating transparency about who accessed what and when.
  - Provides compliance with the legislation and governance frameworks applicable to the ecosystem.

  - Emphasis on transparency and accountability in the ecosystem, generating trust and mitigating risks of misuse or unintended use.
  - Enforces agreements between all participants; provides tools for logging, collects data from all participants and makes use of it.

4.2 Service implementation scenarios for city context

PDM-based services have a number of usage scenarios based on the number of actors in each role. These implementation scenarios and their approach are described in the following table. NOTE: approach describes the current scope and logic for implementing the scenario. Potential future needs and extended capabilities will be considered when the MIM4 specifications evolve further.

Hypothesis:

- Data Source has one connector implementation towards all potential operators, i.e., the same connectivity solution is used for all cases.
- A particular service end-user uses always the same service and operator, i.e., service for a particular user does not connect via multiple instances / operators to the Data Source(s).
- All shareable data (i.e., accessed by services beyond internal entity such as city internal personal data like employment) are implemented with an access gateway -compatible standardized connectivity / service management implementation.
- Same service can be executed as multiple service instances, each having its own configuration, i.e. connected MyData Operator and Data Sources.
- A particular service instance accesses the data directly from one or more Data Sources, but it requires permission from a single MyData Operator, which manages permissions, settlements etc. towards the Data Sources.
- Initially each city has its own governance model and/or rulebook defining the contracts and principles within one operator’s domain. Collaboration between operators will be defined in the MyData Operator network rulebook available later.

#	Scenario	Service	Operator	Data Source	Description / approach
1.	Service using one operator and one	1	1	1	Base case. Supported by reference implementation.
2.	Service using one operator connected to multiple Data Sources.	1	1	N	Operator converts service requests to corresponding Data Source and manages identity mapping between service and Data Sources. Permission for required Data Sources are justified individually but managed as an atomic operation, i.e.,

					permission is valid only if user permits access to all required Data Sources.
3.	Multiple services use the same Data Source over the same operator	N	1	1	Base case (1.) extension. Each service has its own configuration and agreement with a single operator and Data Sources. Operator manages permissions and data settlements for that case.
4A.	Multiple service instances of the same service use different operators to access the same Data Source.	N	N	1	Service provider runs multiple instances of the same core service e.g., for different cities and their operators. Each service instance has its own agreement with a single operator and Data Sources, unless operator buys bulk access to the Data Source over all services. Data Source uses same connector / access gateway-implementation, but agrees separately with each operator the collaboration principles (e.g., governance, trust, settlements).
4B.	One or more services / service instances use different operators to access multiple Data Sources.	1	N	N	Combination of 2 and 4A. Each instance is governed and executed separately.
5.	One Data Source is used in multiple services over multiple operators	N	N	1	Version of 4A. Data Source agrees with the operator the trust and settlement mechanisms for each service. Settlement for a single service is conducted via multiple operators. This scenario requires a common agreement and governance mechanism to be supported and is currently not implemented.

4.3 Exemplary processes for personal data management

The MyData declaration and its associated guiding principles provide a good starting point for defining the key processes and capabilities needed for personal data management for cities. In this chapter we will introduce a set of common processes related to creation, execution, and administration of personal data related services in the smart city environment. At the heart of these processes is trusted collaboration between four roles involved in MyData principles: the Data Using Service (MyData service), MyData Operator, Data Source and Individual (person).

4.3.1 Service Creation

Related user stories	As a service developer, I want to develop a personal data-based service and connect it to a MyData Operator. As a service developer, I want to find relevant external personal Data Sources that are using my connected MyData Operator for permission management. As an owner of a MyData service, I want to identify and document the personal data usage aspects and ensure that those match with the agreed governance principles set by the owners of my identified Data Sources. As a service developer, I want to use the permission management workflow of my connected MyData Operator to communicate my Data Source related personal data requests to persons of my interest. As an owner of a MyData service, I want to have clear human and machine-readable data sharing agreements to ensure compliance for each data transfer. As a Data Source, I want to publish MyData usage policies, requirements, governance so that the permission management can take those into account when allowing access to data to MyData services. As a MyData Operator or MyData service, I want to publish and make available information about my organisation's, policies, certifications and so on so as to prove to Data Sources, I respect their requirements.
Federation over cities	As a service developer, I want to find relevant and external personal Data Sources associated with any of the MyData Operators that my connected MyData Operator has agreed to form a trust group with. As a service developer, I want to use the permission management of my connected MyData Operator when requesting data from any of the external personal Data Sources visible via the trust group. As a service developer, I want to know the data sharing policies of all Data Sources connected to an operator of the trust group and be able to match them with my profile.
Requirements for MIM4	**Current scope** • Define service management and governance principles for collaboration between the Data Source and MyData Operator based on common rulebook and agreed governance principles • Define the data access and permission control between MyData service, MyData Operator, and Data Source • Define the logging, audit trail and validation of access requests between MyData Operator and Data Source **Future scope** • Support multiple types of personal Data Sources beyond currently implemented centralised data registries (e.g., SOLID, decentralised Data Sources, etc.) • Define federated permission management between multiple operators.

<table>
<tr><td></td><td colspan="9">• Define federated audit trail and logging between multiple operators.
• Define standardized, machine readable data sharing agreements and data policies.</td></tr>
<tr><td>MIM4 Actors</td><td colspan="9">MyData service owner
Service developer
MyData Operator(s)
Data Source</td></tr>
<tr><td rowspan="2">MyData and MIM4 linkage</td><td>Identity management</td><td>Permission management</td><td>Service management</td><td>Value Exchange</td><td>Data model management</td><td>Personal data transfer</td><td>Personal data storage</td><td>Governance support</td><td>Logging and accountability</td></tr>
<tr><td>X</td><td>X</td><td>X</td><td>X</td><td>X</td><td>X</td><td></td><td>X</td><td></td></tr>
</table>

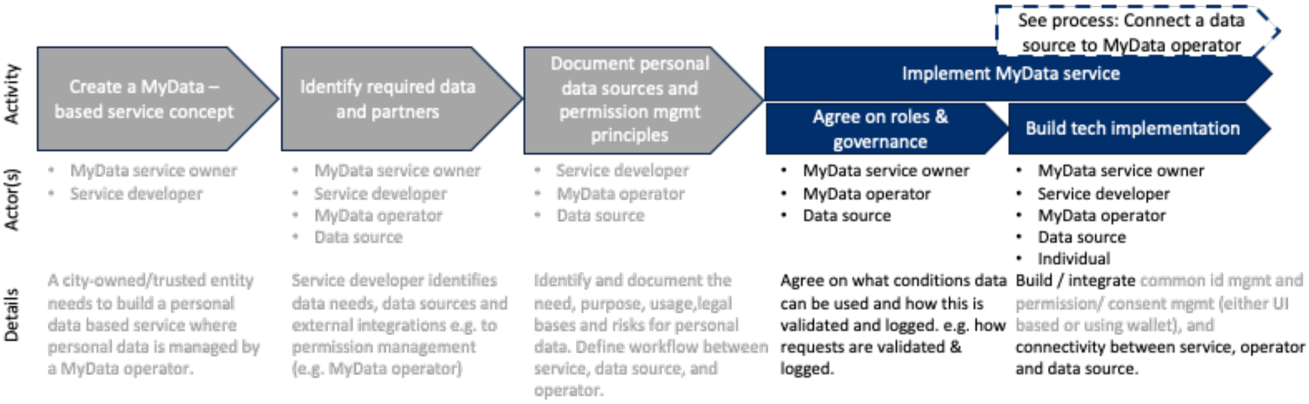

Activity
Actor(s)
Details
Create a MyData – based service concept
Identify required data and partners
Document personal data sources and permission mgmt principles
See process: Connect a data source to MyData operator
Implement MyData service
Agree on roles & governance
Build tech implementation
• MyData service owner
• Service developer
• MyData service owner
• Service developer
• MyData operator
• Data source
• Service developer
• MyData operator
• Data source
• MyData service owner
• MyData operator
• Data source
• MyData service owner
• Service developer
• MyData operator
• Data source
• Individual
A city-owned/trusted entity needs to build a personal data based service where personal data is managed by a MyData operator.
Service developer identifies data needs, data sources and external integrations e.g. to permission management (e.g. MyData operator)
Identify and document the need, purpose, usage,legal bases and risks for personal data. Define workflow between service, data source, and operator.
Agree on what conditions data can be used and how this is validated and logged. e.g. how requests are validated & logged.
Build / integrate common id mgmt and permission/ consent mgmt (either UI based or using wallet), and connectivity between service, operator and data source.

Figure 8: Key processes – MyData-based service creation - MIM4 scope

4.3.2 Service Execution

Related user stories	As a MyData service, I want to identify the Individual as well as ask and retrieve a permission to access personal data behind a Data Source from an Individual by using my MyData Operator service. As a MyData service, I want to retrieve data (for which I have a valid permission) from the Data Source. As a MyData service, if I find out that I didn't gain Individual's permission to retrieve personal data from the Data Source yet and the request was not declined by the Individual, I may re-ask the Individual to provide a permission by using my MyData Operator's service. As a Data Source I want to validate via my MyData Operator that an incoming data request is valid. As a Data Source or Data Using Service or person, I want each data transfer to be linked to a data sharing agreement and data using policy that can be tracked and I

<table>
<tr><td></td><td colspan="9">can find all agreements that concern me in a central place, no matter which operator was used.
As a Data Source I want to have personal data processing related audit trail at hand, on a per MyData service and Individual basis.</td></tr>
<tr><td>Federation over cities</td><td colspan="9">As a MyData service, I want to identify the Individual as well as ask and retrieve a permission from an Individual to access personal data behind several Data Sources not directly connected to my MyData Operator by using my MyData Operator’s service.
As a MyData service, if I don’t have a data sharing agreement with a Data Source that is not connected to my operator but to another in the trust group, I want to be able to generate a data sharing agreement based on the policies and templates of this Data Source through my operator.
As a MyData service, if I find out that I didn’t gain Individual’s permission to retrieve personal data from all Data Sources yet and a particular request was not declined by the Individual, I may re-ask the Individual to provide a permission to this Data Source by using my MyData Operator’s service.
As a Data Source I want to have personal data processing related audit trail at hand, on a per MyData service, per MyData Operator used for permission management, and per Individual basis.</td></tr>
<tr><td>Requirements for MIM4</td><td colspan="9"><b>Current scope</b>
• Agree on what conditions data can be used and how this is validated and logged. e.g., how requests are processed between the MyData Operator and Data Source.
<b>Future scope</b>
• Define required federation and roaming (e.g., id federation, permission management, data sharing policies) between multiple Data Sources and/or MyData Operators
• Define federated governance principles between entities under separate rule books / governance models</td></tr>
<tr><td>MIM4 actors</td><td colspan="9">MyData service
MyData Operator(s)
Data Source</td></tr>
<tr><td rowspan="2">MyData and MIM4 linkage</td><td>Identity management</td><td>Permission management</td><td>Service management</td><td>Value Exchange</td><td>Data model management</td><td>Personal data transfer</td><td>Personal data storage</td><td>Governance support</td><td>Logging and accountability</td></tr>
<tr><td>X</td><td>X</td><td>X</td><td></td><td></td><td>X</td><td>X</td><td>X</td><td>X</td></tr>
</table>

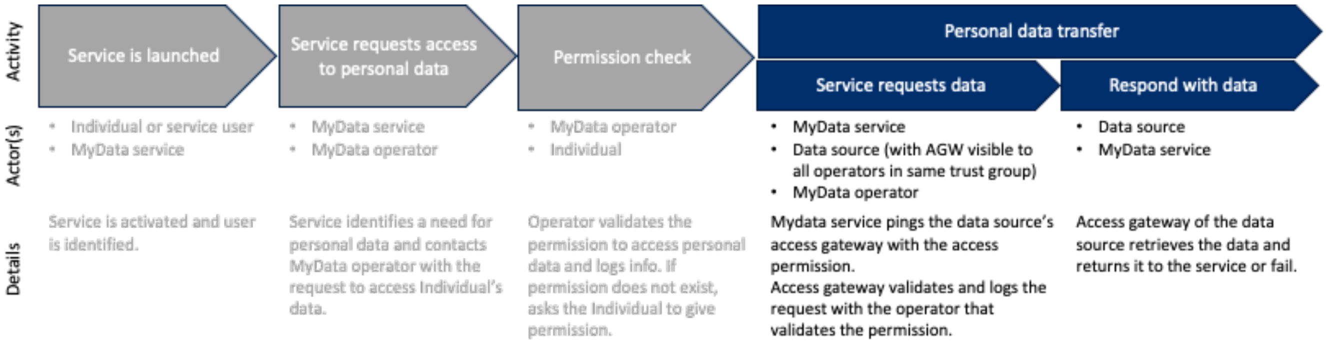

Activity
Actor(s)
Details
Service is launched
Service requests access to personal data
Permission check
Personal data transfer
Service requests data
Respond with data
• Individual or service user
• MyData service
• MyData service
• MyData operator
• MyData operator
• Individual
• MyData service
• Data source (with AGW visible to all operators in same trust group)
• MyData operator
• Data source
• MyData service
Service is activated and user is identified.
Service identifies a need for personal data and contacts MyData operator with the request to access Individual's data.
Operator validates the permission to access personal data and logs info. If permission does not exist, asks the Individual to give permission.
Mydata service pings the data source's access gateway with the access permission.
Access gateway validates and logs the request with the operator that validates the permission.
Access gateway of the data source retrieves the data and returns it to the service or fail.

Figure 9: Key processes – MyData-based service execution - MIM4 scope

4.3.3 Permission and usage control

Related user stories	As the MyData Operator of the MyData service, I will provide the privacy notice towards Individual for approval/denial/information. As a MyData Operator, I want to monitor and track personal data usage and report the use to the Individual. As an Individual, I can either accept or decline a privacy notice that relates to personal data processing under consent or contract. Processing may require personal data transfer from a Data Source/multiple Data Sources to the requesting MyData service. As an Individual, I get notified by my own MyData Operator when personal data processing related to me happens at MyData service under GDPR legal basis other than consent or contract. As an Individual, I can browse actions and actors related to my personal data independent of the legal basis for using it (i.e. also for other permissions than consent or contract). As a Data Source, I want the data sharing agreements I have (or not) to be checked by the permission management to enable the data sharing or not. As a Data Using Service or MyData Operator, I am able to publish my profile (geography, certifications, compliance, data processing, security, etc.) so as to prove I meet the requirements to access Data Sources.
Actors	MyData Operator Individual or service user
Federation over cities	As a designated MyData Operator of the MyData service, I will provide the privacy notice towards Individual for her *approval/denial/information* even if the request originates from another MyData Operator (future scope). As an Individual, I can either accept or decline a privacy notice that relates to personal data processing under consent or contract. Processing may require personal data transfer from a Data Source/multiple Data Sources to the requesting MyData service.

<table>
<tr><td></td><td colspan="9">As a primary MyData Operator for an Individual, I want to monitor and track personal data usage and report the use by collecting usage logs from all related to MyData Operators (usage federation).
As an Individual, my operator can interact with all Data Sources and MyData services connected to the trust group to access data sharing agreements that concern me and review them or generate appropriate consents.
As an Individual, I can either accept or decline a privacy notice that relates to personal data processing under consent or contract. If the request originates from other MyData Operator than my primary operator, I can federate the request to that operator via my local operator (future scope). Processing may require personal data transfer from multiple Data Sources to the requesting MyData service.
As an Individual, I get notified by my own MyData Operator, potentially via federation with other operators (future scope), when personal data processing related to me happens at MyData service under GDPR legal basis other than consent or contract. Processing may require personal data transfer from multiple Data Sources to MyData service.
As a Data Source I want to validate, that an incoming data request from a particular MyData service can be handled by the group of MyData Operators I am currently connected with.
As a Data Source I want to validate via any MyData Operator I connect with, that an incoming data request from a MyData service has a valid permission independent of which operator the request originated from (future scope).</td></tr>
<tr><td>Requirements for MIM4</td><td colspan="9"><b>Current scope</b>
• NONE
<b>Future scope</b>
• NONE</td></tr>
<tr><td rowspan="2">MyData and MIM4 linkage</td><td>Identity management</td><td>Permission management</td><td>Service management</td><td>Value Exchange</td><td>Data model management</td><td>Personal data transfer</td><td>Personal data storage</td><td>Governance support</td><td>Logging and accountability</td></tr>
<tr><td>X</td><td>X</td><td>X</td><td></td><td></td><td></td><td></td><td>X</td><td>X</td></tr>
</table>

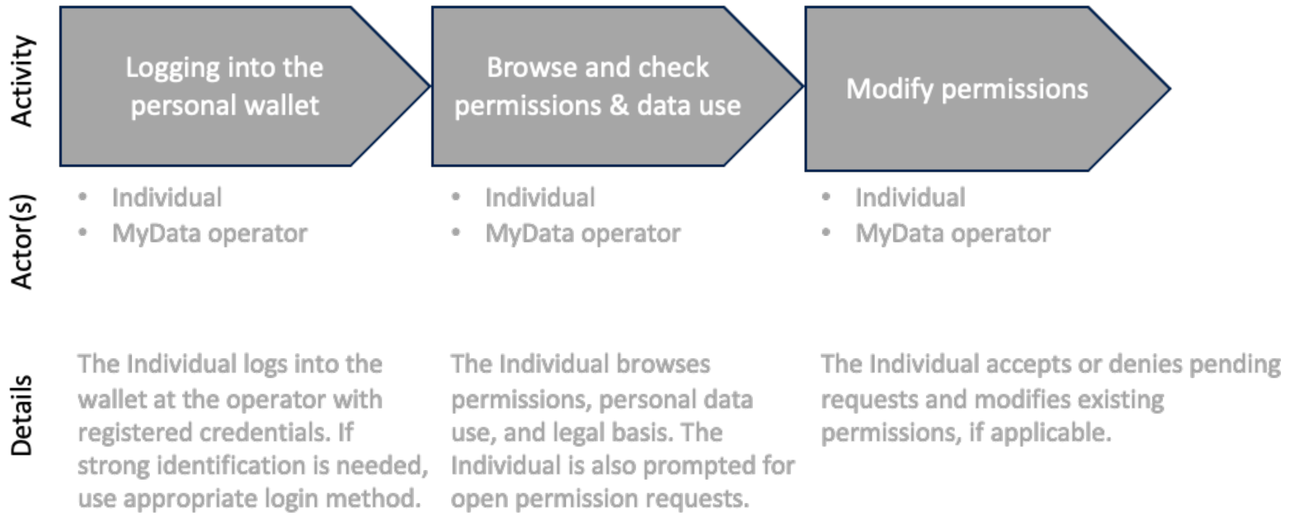

Activity
Actor(s)
Details
Logging into the personal wallet
Browse and check permissions & data use
Modify permissions
• Individual
• MyData operator
• Individual
• MyData operator
• Individual
• MyData operator
The Individual logs into the wallet at the operator with registered credentials. If strong identification is needed, use appropriate login method.
The Individual browses permissions, personal data use, and legal basis. The Individual is also prompted for open permission requests.
The Individual accepts or denies pending requests and modifies existing permissions, if applicable.

Figure 10: Key processes – Permissions and usage control – NO MIM4 Impact

4.3.4 Data usage monitoring and settlement

<table>
<tr><td>Related user stories</td><td colspan="9">As a MyData Operator, I want to track and settle the transactions with MyData service and Data Source according to the contracts and agreements between parties as defined in the common rulebook or in specific/bilateral data sharing agreements.</td></tr>
<tr><td>Actors</td><td colspan="9">MyData service
MyData Operator
Data Source</td></tr>
<tr><td>Federation over cities</td><td colspan="9">As a primary MyData Operator, I want to settle the transactions with MyData service and Data Source according to the contracts and agreements between parties as defined in the common rulebook. For this, I'll collect the usage information from all relevant MyData Operators and conduct other relevant activities (future scope).</td></tr>
<tr><td>Requirements for MIM4</td><td colspan="9"><b>Current scope</b>
• NONE
<b>Future scope</b>
• Interaction with the audit trail and relevant data sharing agreements</td></tr>
<tr><td rowspan="2">MyData linkage</td><td>Identity management</td><td>Permission management</td><td>Service management</td><td>Value Exchange</td><td>Data model management</td><td>Personal data transfer</td><td>Personal data storage</td><td>Governance support</td><td>Logging and accountability</td></tr>
<tr><td></td><td></td><td>X</td><td>X</td><td></td><td></td><td></td><td>X</td><td>X</td></tr>
</table>

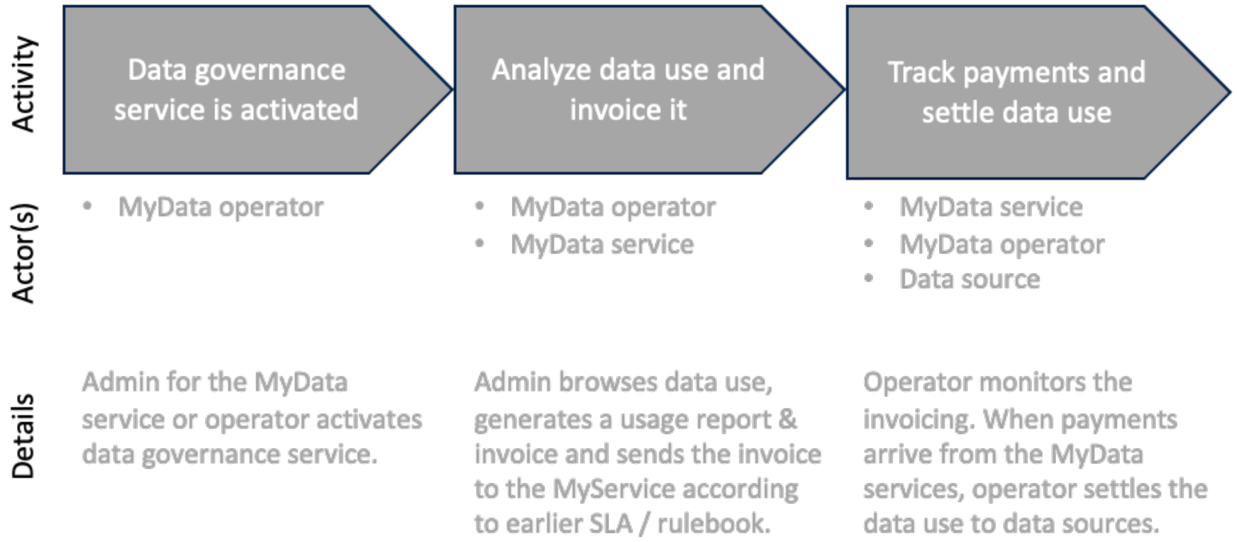

Activity
Actor(s)
Details
Data governance service is activated
Analyze data use and invoice it
Track payments and settle data use
• MyData operator
• MyData operator
• MyData service
• MyData service
• MyData operator
• Data source
Admin for the MyData service or operator activates data governance service.
Admin browses data use, generates a usage report & invoice and sends the invoice to the MyService according to earlier SLA / rulebook.
Operator monitors the invoicing. When payments arrive from the MyData services, operator settles the data use to data sources.

Figure 11: Key processes – Data usage monitoring and settlement – No MIM4 impact

4.3.5 Connect a new Data Source to a MyData Operator

Related user stories	As a MyData service, I need access to a personal Data Source so that permissions are managed via a MyData Operator. As a MyData Operator, I need to connect the Data Source to my services for permissions and settlements. We need also agree on the id federation, governance and settlement principles as defined by our common rulebook. As a Data Source, I need to implement a connectivity solution to connect my Data Source with the MyData Operator and agree on the various id federation, governance and settlement principles between the Data Source and the MyData Operator. NOTE: This process assumes that the same connectivity implementation is used at the Data Source for all MyData Operators.
Federation over cities	As a MyData service, I need access to a personal Data Source available so that permissions are managed via a MyData Operator. As a MyData Operator, I need to connect the Data Source to my services for permissions and settlements. The Data Source might already be connected to some other MyData Operator and is potentially a member of some other rule book, in which case we need to define common rules between networked parties (future scope). As a Data Source, I already have already implemented the connectivity solution with a MyData Operator. If this request requires federation with a new MyData Operator, we need to configure the connectivity solution as well as to agree the various id federation, governance and settlement principles between the Data Source and the operator so that we can collaborate with all connected MyData

<table>
<tr><td></td><td colspan="9">Operators (e.g. by signing the rule book related to the new operator or a common network operator rulebook).</td></tr>
<tr><td>Requirements for MIM4</td><td colspan="9"><b>Current scope</b>
• Define service management and governance principles for collaboration between the Data Source and MyData Operator based on common rule book and agreed governance principles.
• Agree on what conditions data can be used and how this is validated and logged. E.g., how requests are processed between the MyData Operator and Data Source.
• Implement connectivity between MyData Operator and Data Source.
<b>Future scope</b>
• Define settlement principles between Data Source and multiple MyData Operators.
• Federate connectivity, id federation, and permission management between multiple MyData Operators.
• Define federated governance principles between entities under separate rule books / governance models.
• Federation of the data sharing policies set by the Data Source across different MyData Operators.
NOTE: This process assumes that the same connectivity implementation is used at the Data Source for all MyData Operators.</td></tr>
<tr><td>MIM4 Actors</td><td colspan="9">MyData Operator
Data Source</td></tr>
<tr><td rowspan="2">MyData and MIM4 linkage</td><td>Identity management</td><td>Permission management</td><td>Service management</td><td>Value Exchange</td><td>Data model management</td><td>Personal data transfer</td><td>Personal data storage</td><td>Governance support</td><td>Logging and accountability</td></tr>
<tr><td>X</td><td></td><td>X</td><td>X</td><td>X</td><td>X</td><td></td><td>X</td><td>X</td></tr>
</table>

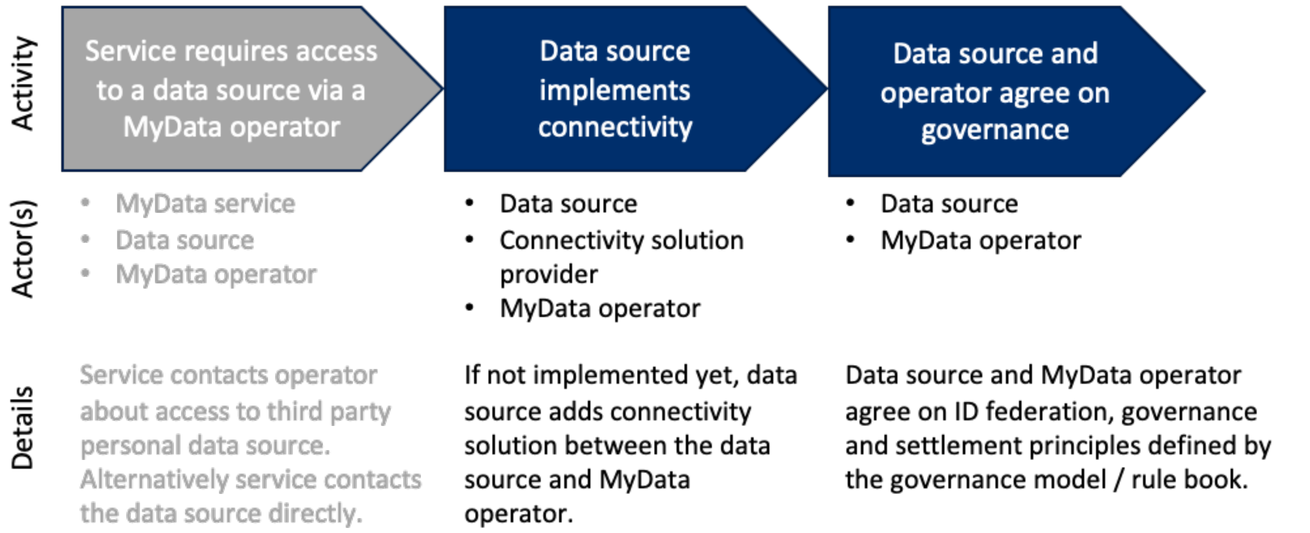
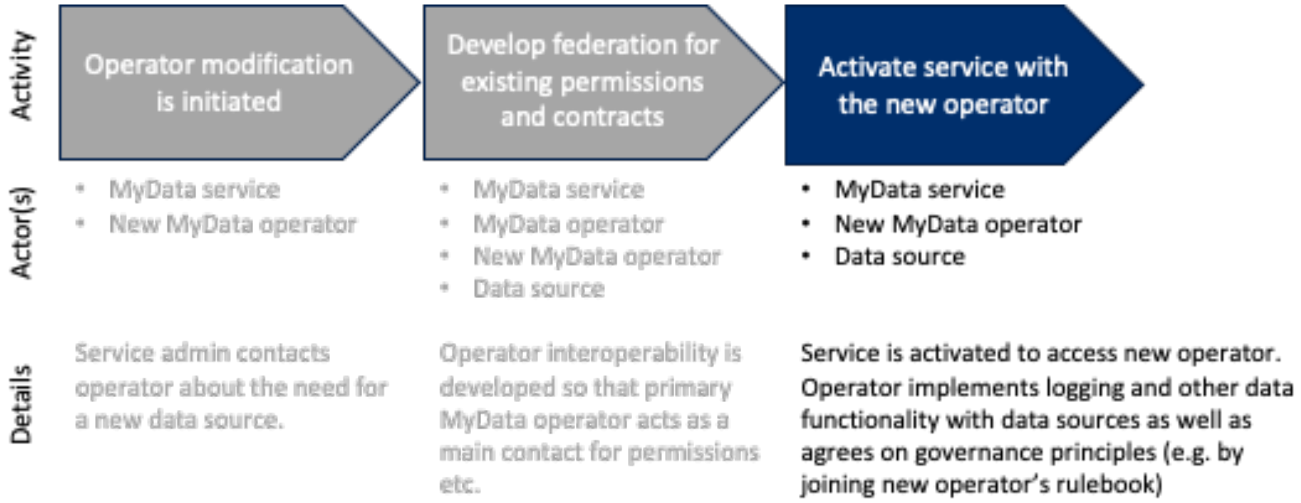

Activity
Service requires access to a data source via a MyData operator
Data source implements connectivity
Data source and operator agree on governance
Actor(s)
MyData service
Data source
MyData operator
Data source
Connectivity solution provider
MyData operator
Data source
MyData operator
Details
Service contacts operator about access to third party personal data source. Alternatively service contacts the data source directly.
If not implemented yet, data source adds connectivity solution between the data source and MyData operator.
Data source and MyData operator agree on ID federation, governance and settlement principles defined by the governance model / rule book.

Figure 12: Key processes – Connecting a new Data Source to a MyData Operator - MIM4 scope

4.3.6 Changing the MyData Operator for a Data Using Service

Related user stories	As a MyData service, I need to provide the service using other operator solution than my current operator connecting to the same Data Source. As a new MyData Operator, I need to agree with the Data Source for the data use as well as agree with the existing MyData Operator on the specifics on the federation between operators (e.g., ID federation, permission, settlements, user interaction, data sharing policies) and connect to the Data Source. As a Data Source, I already have implemented the connectivity solution with a MyData Operator. Now we need to configure the connectivity solution for the new MyData Operator as well as to agree the various id federation, governance and settlement principles between the Data Source and the operator so that we can collaborate with all connected MyData Operators (e.g. by signing the rule book related to the new operator).
Federation over cities	No changes to the user stories as they already apply to city federation as well.
Requirements for MIM4	**Current scope** • Define service management and governance principles for collaboration between the Data Source and the new MyData Operator based on common rule book and agreed governance principles. • Agree on what conditions data can be used and how this is validated and logged with the new operator. • Integrate new MyData Operator to the connectivity solution for the Data Source. **Future scope**

<table>
<tr><td></td><td colspan="9"><ul><li>Federate connectivity, id federation, data sharing policies and permission management between multiple MyData Operators using a primary MyData Operator.</li><li>Define federated governance principles between entities under separate rule books / governance models (e.g., by signing the rule book related to the new operator).</li></ul></td></tr>
<tr><td>MIM4 actors</td><td colspan="9">MyData service
MyData Operator
Data Source</td></tr>
<tr><td rowspan="2">MyData and MIM4 linkage</td><td>Identity management</td><td>Permission management</td><td>Service management</td><td>Value Exchange</td><td>Data model management</td><td>Personal data transfer</td><td>Personal data storage</td><td>Governance support</td><td>Logging and accountability</td></tr>
<tr><td>X</td><td>X</td><td>X</td><td>X</td><td>X</td><td></td><td></td><td>X</td><td>X</td></tr>
</table>



Service admin contacts operator about the need for a new data source.

Operator interoperability is developed so that primary MyData operator acts as a main contact for permissions etc.

Service is activated to access new operator. Operator implements logging and other data functionality with data sources as well as agrees on governance principles (e.g. by joining new operator's rulebook)

Figure 13: Key processes – Changing the MyData Operator for a Data Using Service - MIM4 scope

5 Governance Principles

Governance frameworks, contracts, and data sharing agreements are needed for organisations in the network to trust each other and diverse data intermediaries. Each organisation requires assurance that data transfers are compliant and respect the mutually agreed rules (regulations, sectoral rules, code of conduct contracts, etc). MyData Operators need to have unambiguous contracts between themselves, people, as well as organisations.

From data governance perspective, the MIM4 solution provides a legal trust framework, that is integrated natively to the technical use of MIM4, such as MIM4 connectivity specification. This work builds on on-going initiatives such as Sitra Fair Data Economy Rulebook model and the **aNewGovernance** (aNG) initiative funded by the European Commission. Both initiatives aim to augment the MIM4 connector with a robust supporting legal framework.

5.1 Ecosystem Governance

MyData principles provide a recognized and solid ground for MIM4 related to personal data management. Critical layer is the management of data sharing agreements. It should cover GDPR / consent, contract, legal obligation, vital interests of the data subject, public interest, and other legitimate reasons.

Examples of legislation governing personal data ecosystems in the context of cities include:

- GDPR and similar personal data protection frameworks around the world
- Sector-specific regulations, such as health and financial sector
- EU's data specific effort, such as EU Data Strategy. Data Governance Act, or upcoming Data Act legislation

As recognized in the Proposal for a Regulation on European data governance (Data Governance Act), providers of data sharing services (data intermediaries) are expected to play a key role in the data economy, as a tool to facilitate the aggregation and exchange of substantial amounts of relevant data. Data intermediaries offering services that connect the different actors have the potential to contribute to the efficient pooling of data as well as to the facilitation of bilateral data sharing. Specialised data intermediaries that are independent from both Data Providers and End-Users can have a facilitating role in the emergence of new data-driven ecosystems independent from any player with a significant degree of market power.[18] To comply with the Data Governance Act, its provisions will be considered, when they will become available, while this Minimum Interoperability Mechanism is developed further.

In order to produce a vast personal and human-centric data sharing infrastructure there **needs to be a precise legal framework** with a contractual framework for data sharing agreements taking into account **industry specific regulations, ethics charters, codes of conduct and determining each party's responsibilities** and liabilities in the data transfer. This legal model needs to be backed by a strong **legal infrastructure** to be able to automate the generation, signature, auditability, enforcement and traceability of these contracts and governance frameworks. This legal infrastructure should be the **collective work** of parties contributing to it and benefiting from it.

In the future evolution of the MIM4 -specifications, we propose to precise the layer 2 of MIM4 (Data sharing agreements) to build an **open legal infrastructure** to automate the generation of **governance**,

[18] Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) COM(2020) 767 final.

contracts and agreements for data sharing between the person, the MyData Operator, the Data Sources and the Data Using Services.

This legal infrastructure will allow Data Sources to set data sharing policies (obligations, restrictions, requirements, certifications, pricing, etc), data services to be matched or not with such policies and data sharing agreements to be generated, notarized, machine readable and tokenized to enable the data exchange. This legal infrastructure should enable interoperability of data sharing agreements across Data Sources, data services, MyData Operators of a trust group, to enable the user stories described earlier.

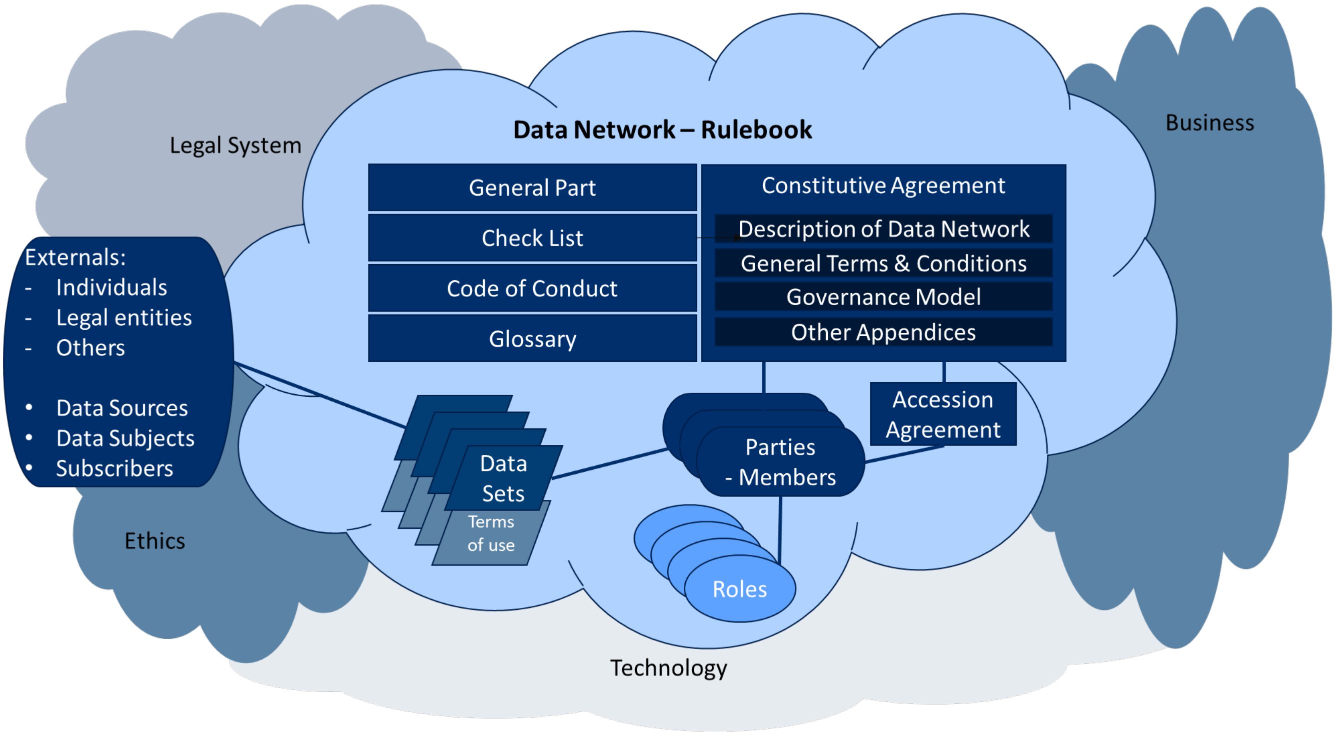
This will augment the trust, auditability and traceability of each data transfer **providing strong human-centric governance** of the infrastructure and can easily be scaled to more members and other sectors.

To build such a contractual framework between Data Sources, Data Using Services and MyData Operators, the Sitra Rulebook is the most relevant basis.

5.2 Sitra Rulebook Template for Data Networks

The purpose of the Sitra Rulebook Template is to provide an easily accessible manual on how to establish a data network and to set out general terms and conditions for data sharing agreements. The Rulebook Template helps organisations to form new data networks, implement rulebooks for those data networks, and promote the fair data economy in general.

The following picture illustrates the relations between the different parts of a rulebook implemented from the Rulebook Template.



Data networks that adopt the Sitra Rulebook Template must be fair, balanced, and lawful in their processing of data. They must also be just and impartial toward their members and ensure that the rights of third parties are not infringed. Personal data must be processed in accordance with European and applicable national data protection regulations. Data networks identify and manage risks associated with the sharing and processing of data while ensuring the exploitation of new possibilities that data offers. This includes also ensuring compliance with relevant competition legislation and that the data network will not have a negative impact on market competition and consumers. Provisions restricting access to the network are especially important to consider in this kind of assessment.

It is essential to realize that each of the members of the data network can operate in several roles and they may change continuously. Service Providers and Infrastructure Operators are natural candidates for independent data intermediaries in accordance with the Data Governance Act. Nonetheless, it should be noted that they are not always independent, but Data Providers and End-Users may occasionally also provide services or operate the infrastructure. Also, note that in a wider context, even Data Providers may get data from external sources and there can be external parties, Subscribers, that receive data from the data network in accordance with the Data Sets' Terms of Use although they are not Parties of the Constitutive Agreement.

5.3 MyData Rulebook for Cities

Based on the generic Sitra Rulebook Template model, a more specific MyData rulebook for cities template entitled **Helsinki Trust Network Rulebook** is under development as part of the MyData operator project for City of Helsinki.

Helsinki Trust Network Rulebook makes public the principles of Personal Data Management applied by the city, and concretely helps different parties to define common functional and technical procedures related to personal data. In addition, it offers model legal agreements for the data network that is exchanging personal data in the city context. These model agreements have been directly derived from the Sitra Rulebook Template to be more applicable for the needs of municipalities and their data networks. However, the changes needed vis-à-vis the generic Sitra Rulebook Template agreements have been relatively minor.

The rulebook also defines principles for the movement of urban data between different cities with the consent of the citizen, thus significantly increasing the potential for the utilization of urban data. Sharing urban data between cities, for example, enables a more integrated service offering based on life situations. In many use cases, data must be able to be exchanged between cities, e.g., (1) shared service production organized as an urban joint venture or organization, (2) a city-managed benefit, or (3) a city-managed training and competency information.

The Helsinki Trust Network Rulebook template is connected with this MIM4 specification as it provides a way to agree, based on common principles and MIM4 adherence, more specific practices for any city that is implementing MIM4. It can also act as the legal basis for implementing MIM4, as parties of the data network would need to agree to follow the rulebook of that city.

The current version of the Helsinki Trust Network Rulebook template is available from the authors and a link to the initial version will be added to this document once the template is publicly released.

As an evolution of the MIM4 specification, this Rulebook will be made machine readable, and the specification of the legal infrastructure will be published.

5.4 Governance of MIM4 Specification

This MIM4 specification will be governed as part of official MIM governance structure defined by OASC.[19] MIM4 is one of the MIMs which are developed by the OASC Technology Council and governed by the Council of Cities and the Board of Directors of OASC.

[19] https://oascities.org/governance/

6 Technical Architecture

This section explains on logical level the overall technical architecture for Personal Data Management in context of smart cities.

6.1 High level architectural requirements

The following figure depicts the architecture requirements for common data spaces as defined by the OpenDEI position paper entitled "Design principles for data spaces". These principles are used as the starting point for defining the architectural requirements for the personal data management and are then modified to reflect especially MyData guidance as part of the core principles.

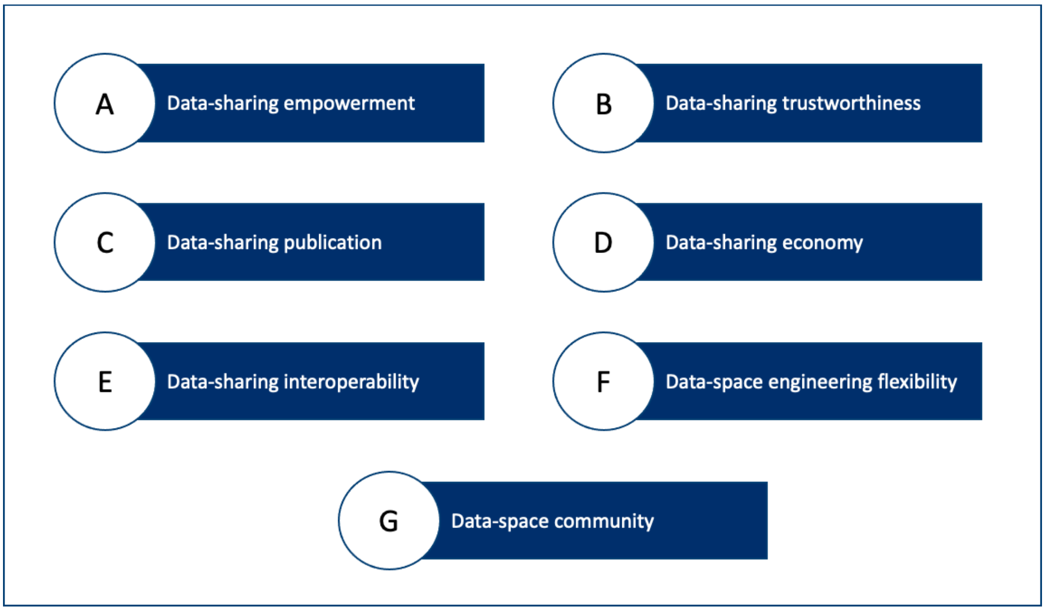


Figure 14: Architecture requirements for data spaces[20]

Requirement A: Data-sharing empowerment is about ensuring that decisions can be made by appropriate stakeholders. This means that tools and organisational practices are available for

- **governance in data spaces**, i.e. the possibility to define and monitor policies in data sharing,
- citizen engagement support, i.e. the possibility for citizens to participate in data sharing and exchange transactions,
- **data sovereignty support**, i.e. the possibility for stakeholders owning data to govern the use of it, and
- **federation**, i.e. the possibility to connect several data platforms with each other, with each one retaining control of its own operations.

[20] Adapted from OpenDEI Position paper: Design principles for data spaces. Available at: https://design-principles-for-data-spaces.org/

Requirement B: Data-sharing trustworthiness is about ensuring that data spaces operate according to expected requirements. This means that the development of data-sharing applications must support

- **security-by-design**, i.e. security of data space assets and support of non-repudiable and unambiguous agreements,
- **privacy-by-design**, i.e. integration of privacy concerns in the development of data platforms and data-sharing applications, and
- **assurance-by-design**, i.e. integration of security and privacy assurance requirements in the development of data platforms and data-sharing applications.

Requirement C: Data-sharing publication is about enabling data to be published so it can be easily located by data consumers.

Requirement D: Data-sharing economy is about creating the conditions for data sharing and exchange, requiring

- **non-financial incentive mechanisms**,
- **financial incentive mechanisms**, including models to monetise data and methods to determine the value of data, and
- **agreement mechanisms**.

Requirement E: Data-sharing interoperability is about providing the ability for all applications in data spaces to create, use, transfer and effectively exchange data. This requires the definition of data exchange APIs and data models supporting

- **semantic interoperability**, ensuring that the meaning of the data model within the context of a subject area is understood by the participating systems,
- **behavioural interoperability**, ensuring that the actual result obtained from usage of data exchange APIs achieves the expected outcome, and
- **policy interoperability**, i.e. interoperability while complying with the legal, organisational, and policy frameworks applicable to the participating systems.

Requirement F: Data space engineering flexibility is about providing the ability for engineers to add customised features in data-processing applications and data platforms to enable flexibility in terms of interoperability, i.e. extending data spaces with specific interoperability capabilities,

- **trustworthiness flexibility**, i.e. extending data spaces with specific security, privacy, and assurance capabilities, and
- **data processing flexibility**, i.e. extending data spaces with data-processing capabilities.

Requirement G: Data space community is about fostering maximum reuse of data space solutions. This includes

- **open solutions**, i.e. ensuring that data space platforms and data-sharing applications are based on open specifications,

- **reusability**, i.e. ensuring that capabilities of existing data and marketplace platforms as well as data-sharing applications can be easily replicated,
- **open source**, i.e. allowing free access to data and marketplace components developed by communities, and
  sustainability of solutions, i.e. assurance that solutions will be available and maintained over a long period of time.

6.2 Conceptual architecture for Personal Data Management

Earlier MIM-related specification effort has developed a generic technical architecture for all MIMs that is heavily geared towards IoT-like use cases. As MIM4 focuses on personal data management and introduces new concepts such as permissions, that architecture has been slightly modified to cover also personal data management perspective as depicted in the following figure.

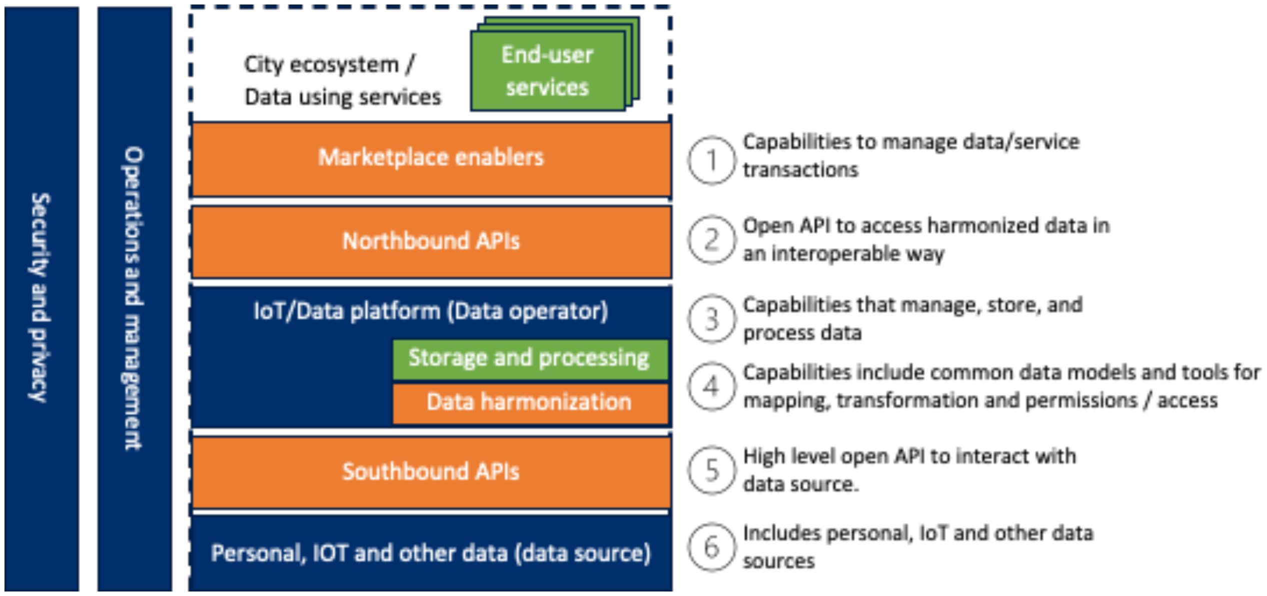

Security and privacy
Operations and management
City ecosystem / Data using services
End-user services
Marketplace enablers
Northbound APIs
IoT/Data platform (Data operator)
Storage and processing
Data harmonization
Southbound APIs
Personal, IOT and other data (data source)
1 Capabilities to manage data/service transactions
2 Open API to access harmonized data in an interoperable way
3 Capabilities that manage, store, and process data
4 Capabilities include common data models and tools for mapping, transformation and permissions / access
5 High level open API to interact with data source.
6 Includes personal, IoT and other data sources

Figure 15: Generic architecture for Minimum Interoperability Mechanisms (MIMs)[21]

When converting the generic architecture into a conceptual architecture for the MIM4 scope, personal data management, we can identify a set of core capabilities that support fair use of personal data among the key stakeholders.

This high-level conceptual architecture is depicted below.

[21] Adapted from https://living-in.eu/group/7/commitments/mims-plus-technical-specifications-v3

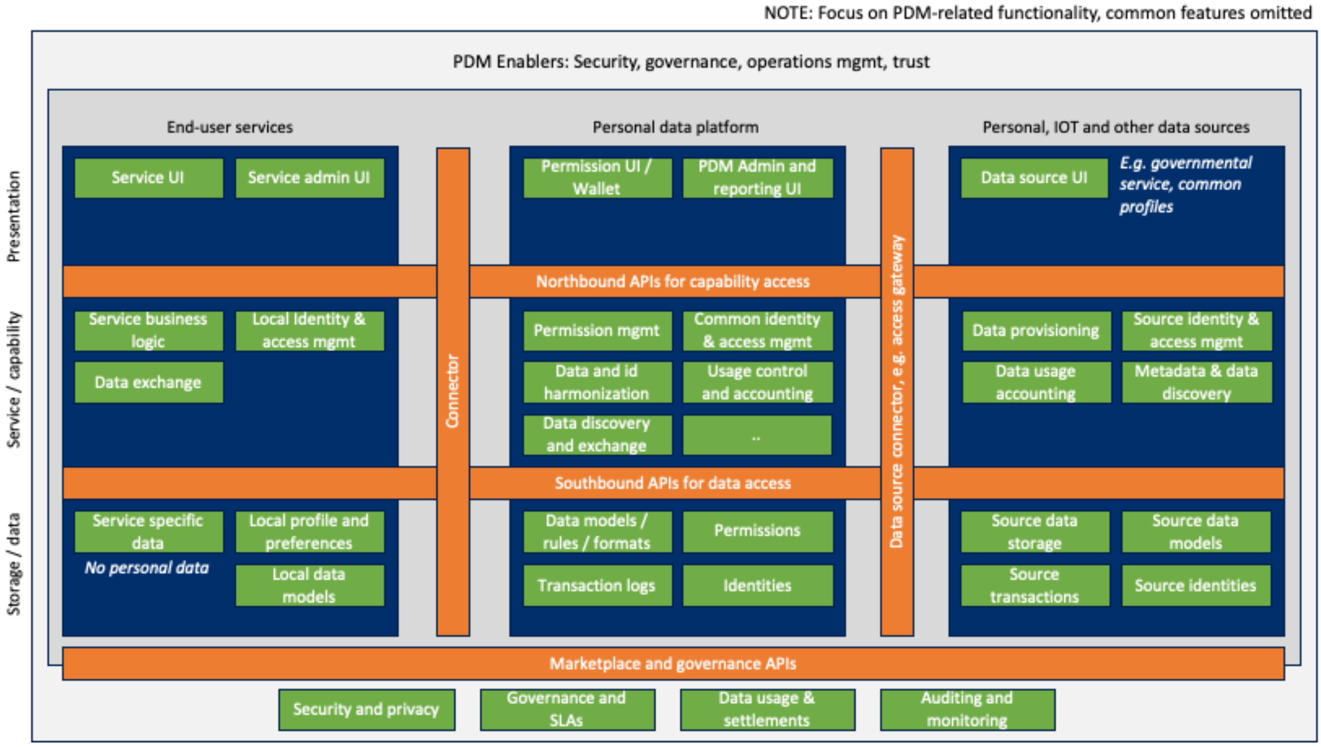

NOTE: Focus on PDM-related functionality, common features omitted
PDM Enablers: Security, governance, operations mgmt, trust
Presentation
Service / capability
Storage / data
End-user services
Service UI
Service admin UI
Service business logic
Local Identity & access mgmt
Data exchange
Service specific data
No personal data
Local profile and preferences
Local data models
Connector
Personal data platform
Permission UI / Wallet
PDM Admin and reporting UI
Northbound APIs for capability access
Permission mgmt
Common identity & access mgmt
Data and id harmonization
Usage control and accounting
Data discovery and exchange
..
Southbound APIs for data access
Data models / rules / formats
Permissions
Transaction logs
Identities
Data source connector, e.g. access gateway
Personal, IOT and other data sources
Data source UI
E.g. governmental service, common profiles
Data provisioning
Source identity & access mgmt
Data usage accounting
Metadata & data discovery
Source data storage
Source data models
Source transactions
Source identities
Marketplace and governance APIs
Security and privacy
Governance and SLAs
Data usage & settlements
Auditing and monitoring

Figure 16: Conceptual architecture for personal data management in city-context

6.2.1 End-user services

End-user services contain a range of services and applications related to personal data. These services provide a user interface for both end users (Individuals) and for administrative use. In addition to service-specific business logic, these services provide typically some form of local identity and access management as well as local manipulation and storage of data.

To be compliant with privacy and data protection legislation and MyData principles, these services must empower the Individual to monitor and control the use of their personal data. This functionality is provided to the end-user services via a common connector. This connector connects the service to a personal data platform managing permissions related to personal data. It also connects the service either indirectly or directly to the actual Data Source. This interface and the underlying personal data management functionality is often provided by a separate entity, MyData Operator.

6.2.2 Personal data platform

Personal data platform acts as a control point between the Data Using Services and Data Sources containing the actual personal data. For Individuals it provides a user interface to monitor and manage the use of personal data in different services. Similarly, for services and Data Sources the personal data platform is a trusted middleman for logging, monitoring, and permitting access to data on behalf of the Individual.

In practice each service request for access to personal data is logged and validated against user-given consents and other legitimate interests. If permission is given, the service can retrieve the personal data using a given permission. When receiving a request for personal data, the Data Source similarly logs the request and checks the validity of the request via the Data Source connector from the personal data platform.

If the service does not have a permission to personal data, the personal data platform can request the Individual to authorize the service to access the personal data for a specific, well-defined need.

Personal data platform may support also other data management related tasks, such as identity mapping between services and Data Sources, data harmonization, as well as settlements and reporting for data usage.

In MyData -supported cases, personal data platform is typically provided by a MyData Operator.

6.2.3 Data Sources

Data Sources contain the personal data needed by the services. Data Sources require the incoming access requests to personal data to contain a permission token, which is then validated by the Data Source connector using MyData Operator capabilities. In practice this functionality is quite often isolated in the connector so that the Data Source provides a generic internal interface for data and trusts the connector to manage personal data related functionality in concert with the MyData Operator.

6.2.4 Interfaces and governing structures

The notion of trust and division of work between different actors are essential for the personal data management. Legal and governance framework, personal data rule books and related contracts provide a foundation for this trust, but these decisions are reflected in the infrastructure capabilities, including e.g., security, privacy, agreements and contracts, and data life cycle capabilities like auditing and monitoring of data.

Currently the infrastructure support for legal and governance aspects is focusing mostly on the security, privacy, trust, and monitoring capabilities.

In the future the underlying infrastructure capabilities will likely evolve towards automating key governance aspects, such as contract lifecycle support as well as enforcement and traceability of the contracts and governance frameworks.

7 Summary and MIM4 connectivity scope

Personal data management is approached in this version of MIM4 specification through two pillars: **Connectivity** and **Legal Framework**. The first pillar is an open-source connector that enables multiple cities to utilise one Data Source without needing to build their own, parallel connectors to this Data Source. The second pillar is a legal framework that governs the use of connectors for data access.

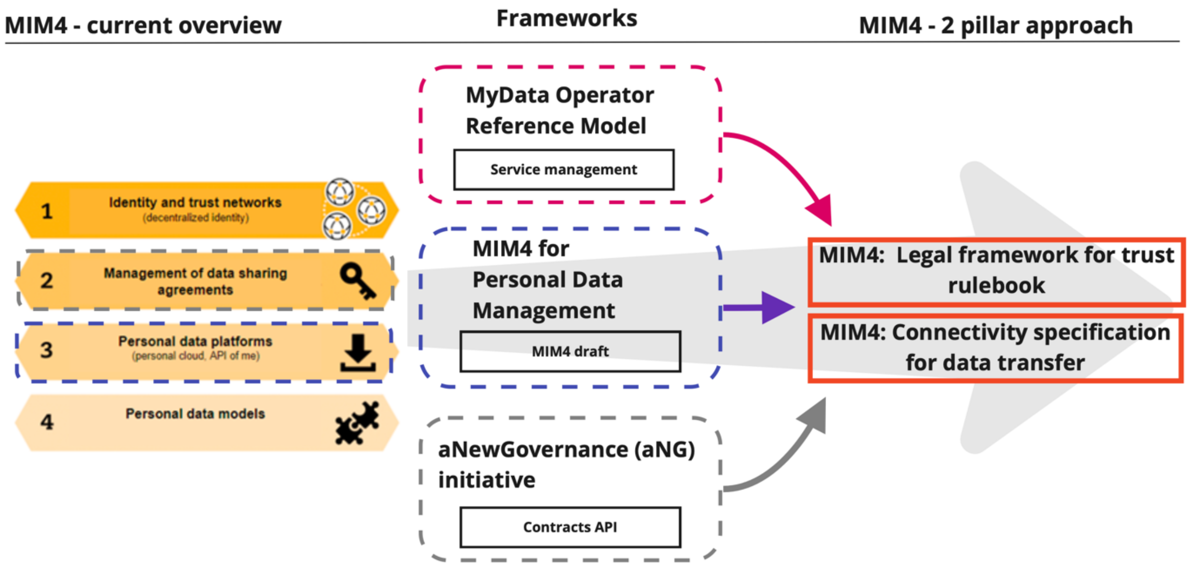

MIM4 - current overview
Frameworks
MIM4 - 2 pillar approach
1
Identity and trust networks
(decentralized identity)
2
Management of data sharing agreements
3
Personal data platforms
(personal cloud, API of me)
4
Personal data models
MyData Operator Reference Model
Service management
MIM4 for Personal Data Management
MIM4 draft
aNewGovernance (aNG) initiative
Contracts API
MIM4: Legal framework for trust rulebook
MIM4: Connectivity specification for data transfer

Figure 17: Relationship of initial MIM4 layers, MyData Operator Reference Model and aNewGovernance data sharing agreement framework.

Pillar 1: MIM4 Connectivity specification for data transfer

The layer 3 Personal data platform APIs of the MIM4 stack offer harmonised data access to MyData Operators. The architectural target is to abstract the operator (permission management services) from the data transfer. This way one Data Source – an existing API or Personal Data Storage - is able to serve data to multiple MyData Operators simultaneously. This capability is envisaged to be required in collaborative arrangements across cities and in use cases where personal data is re-used. Access to a Data Source should be possible with multiple permission management systems and scale to cover future technologies.

As a reference implementation, a **Connector** (i.e., **Access Gateway**) is a freely available and extendable **open-source proxy component** that implements the MIM4 connectivity requirements and allows the same Data Source to be utilised by multiple MyData Operators. Each operator provides their own identity and permission management services and interfaces.

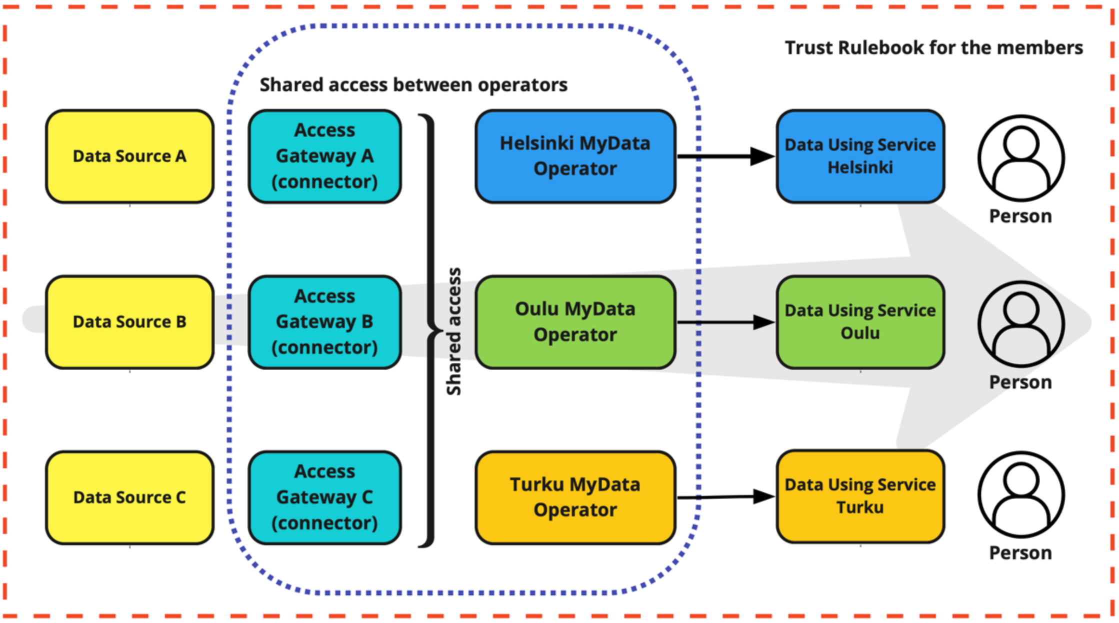


Figure 18: Access Gateway as a shared Connector between multiple MyData Operators and actual Data Source.

Pillar 2: MIM4 Legal Framework for trust rulebook

The legal framework governance is designed to make personal data re-use easier whilst retaining robust permissioning. The legal framework considers each Data Using Service as a customer that requires access to one or more Data Sources. This legal framework is required to validate the status of the permission of the customer request to access the Data Sources via a single point of contact (operator) irrespective of the operator of the Data Source.

The legal framework stipulates trust framework requirements for common rules between operators within a Trust Group (two or more MyData Operators operating in e.g., smart city context). MIM4 legal framework extends the **Helsinki Trust Network Rulebook** for cities to cover data sharing across operators across countries. Multi-city data sharing use cases are handled as inter-operator situations because smart cities working with similar entities may be assuming different roles within the MyData framework. Inter-operator trust frameworks need to be specified and formed.

A well-functioning connectivity layer is necessary but insufficient condition for a multi-operator human-centric network. Governance frameworks, contracts, and data sharing agreements are needed for organisations in the network to trust each other and diverse data intermediaries. Each organisation requires assurance that data transfers are compliant and respect the mutually agreed rules (regulations, sectoral rules, code of conduct contracts, etc). MyData Operators need to have unambiguous contracts between themselves, people, as well as organisations. This work builds on on-going initiatives, such as Sitra Fair Data Economy Rulebook model and the aNewGovernance (aNG), which aim to augment the MIM4 connector with a robust supporting legal framework.

MIM4 Deliverables of the current version

Deliverables of the current MIM4 release include the following

- **Initial version of the MIM4 connectivity specification and future roadmap** (this document)
  - Specification for initial connectivity-layer interoperability
  - Guiding principles for the technical environment and legal ecosystem
- **Use case example from City of Helsinki**
- **An open-source access gateway that facilitates the MIM4 implementation**
  - Initial, extendable version of the MIM4 connector - with interoperability across two or more MyDataShare (Vastuu Group's MyData Operator platform) operator instances
  - Roadmap for multi-vendor support
- **Legal framework and rulebook**
  - Initial Helsinki Trust Network Rulebook for cities
  - Requirements and roadmap for actual multi-operator framework – cross-mapping the MIM4 specification and the Service Management function as defined in the MyData Operator White Paper
  - Requirements and roadmap for machine-readable data sharing agreement management from the aNG initiative

PART 2 - MIM4 –The MIM4 Connectivity Specification

1 Introduction

1.1 Abstract

Part 2 of the MIM4 connectivity specification document defines the key interoperability mechanisms for MyData-based personal data management within city context. The current version of the specification focuses on providing connectivity between personal Data Sources and MyData Operators. In the future the scope will be expanded to cover a larger part of the personal data management processes that are discussed in the Part 1 of the specification.

1.2 Status of this document

This is an initial draft version. Status of the document is a working draft, and it is changing on a daily basis. Please contact the authors for the latest version.

Availability, license and change control mechanisms will be defined later.

1.3 Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119.

1.4 Specification Context and Scope

Part 1 of the specification defines a set of common processes and capabilities for personal data management in the city context. This part focuses on the capabilities currently under the Minimum Interoperability Mechanisms (MIM) 4 – Personal Data Management.

The core business goal through following this MIM4 connectivity specification is: If there exists a trust group of multiple MyData Operators, a Data Using Service that is organising its operation contractually via one such operator shall not need to make bilateral contracts with the other group members to have access to mutually shared Data Sources of the group.

Core functional goal of this connectivity specification, that is Data Source sharing across multiple operators, is presented at logical & functional level in the Figure 19.

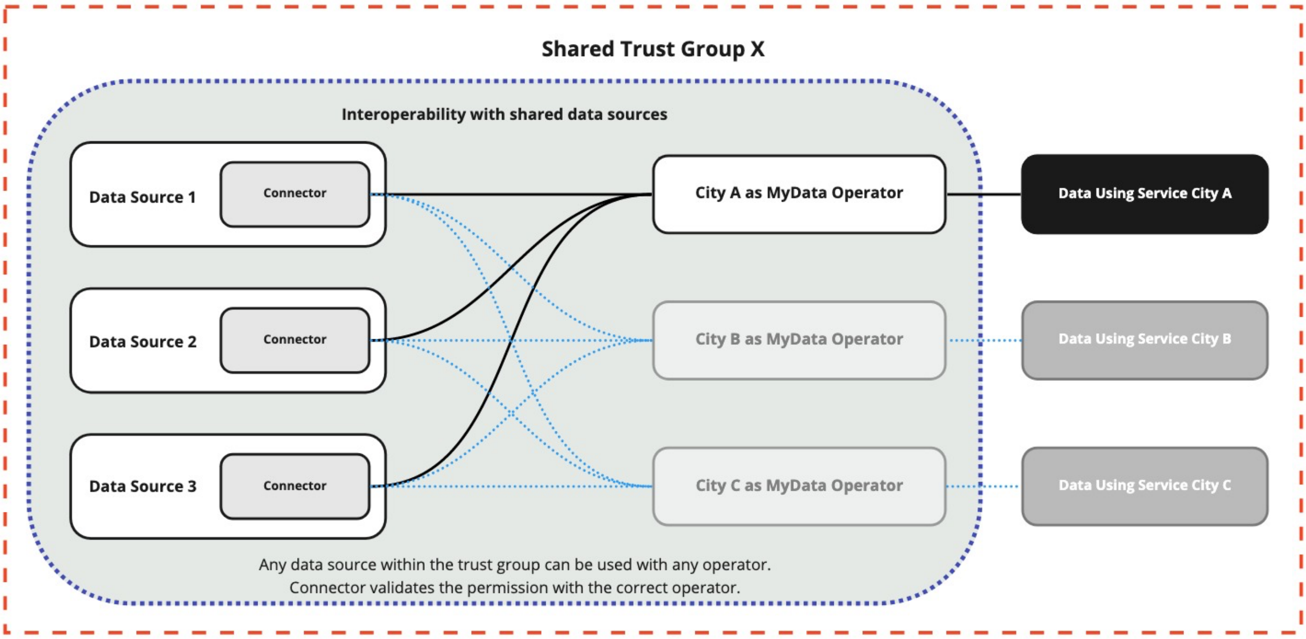

Shared Trust Group X
Interoperability with shared data sources
Data Source 1
Connector
Data Source 2
Connector
Data Source 3
Connector
City A as MyData Operator
City B as MyData Operator
City C as MyData Operator
Data Using Service City A
Data Using Service City B
Data Using Service City C
Any data source within the trust group can be used with any operator.
Connector validates the permission with the correct operator.

Figure 19: Logical components of Data Source sharing across operators

2 Key requirements for MIM4 connectivity

This chapter summarizes the identified key requirements both for the current focus for the specification as well as potential evolution for the specification.

2.1 Requirements for personal data management

Process descriptions in the Part 1 define the context for the MIM4 as a whole. These processes describe generic activities related to personal data management within and between cities.

The initial version of the specification focuses on a subset of these processes. Scope for the initial MIM4 specification is on the connectivity and service management capabilities between MyData Operators and Data Sources.

The following table contains the key requirements for the MIM4 specification divided into two categories, initial requirements and future work. Potential future represents potential evolution for the specification going forward.

ID	Requirement	Actors	Source
A.I	Define service management and governance principles for collaboration between the Data Source and MyData Operator based on common rule book and agreed governance principles	• MyData Operator • Data Source	• Service Creation • Connect a new Data Source to a MyData Operator • Adding a new MyData Operator to a service

A.II	Define the connectivity functionality to be implemented to the Data Source.	• Data Source	• Service Creation
A.III	Define the data access and permission control between MyData service, MyData Operator, and Data Source	• MyData service • MyData Operator • Data Source	• Service Creation
A.IV	Define the logging, audit trail and validation of access requests between MyData Operator and Data Source	• MyData Operator • Data Source	• Service Creation
A.V	Agree on what conditions data can be used and how this is validated and logged.	• MyData service • MyData Operator • Data Source	• Service Execution • Connect a new Data Source to a MyData Operator • Adding a new MyData Operator to a service
A.VI	Implement connectivity between MyData Operator and Data Source.	• MyData Operator • Data Source	• Connect a new Data Source to a MyData Operator
A.VII	Integrate new MyData Operator to the connectivity solution for the Data Source.	• MyData Operator • Data Source	• Adding a new MyData Operator to a service

3 Legal framework, operating principles and governance for MIM4 connectivity

Requirements and roadmap for a multi-operator network are defined in **MyData Operator Network Rulebook**. This rulebook that governs the collaboration between MyData Operators provides a cross-mapping between this MIM4 Connectivity Specification and the Service Management function as defined in the MyData Operator White Paper.

4 Functional specification

4.1 Functional components involved in MIM4 connectivity

Basic architecture and functions of Personal Data Management in context of single MyData Operator were explained in Part 1. This technical specification focuses on enabling sharing of Data Sources across multiple MyData Operators.

Previously covered basic components and roles that are included:

- **Data Source** as the service providing data and as target of personal data re-use. Target with technical realisation of interconnectivity is that existing implementation of the source (API, Personal Data Store or Pod) *doesn't need to be changed from technical perspective*, just accessed through a new client system defined in this document.

- **Data Using Service** - the entity that requires access to one or several Data Sources protected by MyData Operators.
- **Permission Management** which handles access control to personal data provided by Data Sources through executing the chosen technical approach of a particular MyData Operator. The interoperability mechanism SHALL NOT assume a single standard exists for permission management among the trust group members.

The following *new* logical components and roles can be identified in multi-operator environment:

- **Connector** – the central connectivity related component of this specification - an extendable proxy component that implements the MIM4 connectivity baseline technical requirements and provides GDPR-related audit trail of the personal data usage for the regulatorily obligated parties. The connector allows a single Data Source to work in parallel with several MyData Operator platforms' identity and permission management logic and interfaces without technical changes to the original Data Source. Connector also removes the need for carrying personal identifiers (such as GovIDs) across the components, which is a highly undesirable principle from privacy preservation perspective.
- **Trust Group** – a registry function that groups technically together a set of above components. Group contains only the MyData Operators as its members, but helps indirectly to resolve the operators' shared Data Sources and connector information. A trust group and the metadata it includes MUST be made accessible by the connectors to those operators that are part of a trust group. Any component may be simultaneously linked to several Trust Groups.
- A commonly agreed & public **means to query for an up-to-date member list of a Trust Group** (a group of MyData Operators). This means is needed by the connector to validate whether data requests coming from a Data Using Service in a particular operator's domain can be served by the connector. That is, the connector will not just validate the data subject's permission behind the data request is valid, but also that a specific data request is coming *from within a trust group* which the connector resolves via the trust group.

Note

The realisation of practical trust framework within the Trust Group needs to grow from the legal rulebook arrangement between the parties (see Part 2, Section 3). Reflecting the rulebook's section on cross-operator Data Source sharing on the technical level is a public online trust list mechanism whose contents can be queried on real-time. Similar API-level examples to accessing trust lists can be found e.g., from eIDAS[22] context with electronic identification services[23].

Technical (for example, cryptographic signatures or encryption based) implementation of trust across the components is omitted from the first version of the MIM4 specification.

[22] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.257.01.0073.01.ENG

[23] eIDAS Trust List API: https://esignature.ec.europa.eu/efda/swagger-ui.html

Detailed technical descriptions and API endpoint level details of the Connector, Trust Group and trust group management and their functionality with basic Personal Data Management components are specified in Part 2, Section 5.

4.2 High-level relations and functions between components

This section describes how the new connectivity and service management related components relate to a key subset of Personal Data Management functions on the architecture level. Each section explains how the technical implementation supports or carries the functional requirements of a multi-operator environment.

4.2.1 Connector – Data Source relation

An existing Data Source most likely lacks compatibility with any kind of MyData Operator solution. Even if there would be support for a single vendor, the emerging MyData service market has multiple operator solutions competing for customers. The proposed technical solution to foster scalable multi-operator interoperability on the Data Source side is a connector, an easily configurable *proxy service* that can adapt several MyData Operators' technical interfaces and protocols.

Specific functional requirements of this tightly coupled Data Source – Connector relation are:

- As a proxy to the Data Source (a server, assumed in this document) the connector MUST be provided with suitable client configuration and access authorisation secrets towards the Data Source. There exist several solutions for API access management in the market, and such should be pluggable into the proxy service as client library level components by developers.
- As a proxy to the Data Source the connector MUST be configured to parse the inbound request from the Data Using Service to one or several requests towards the Data Source. At its simplest, these inbound and source requests are identical, but the proxy may implement any required parsers, processors (e.g. from one schema or format to another) and response generations that are needed.
- As a proxy to the Data Source the connector SHALL introspect the request performed by the Data Using Service, and only submit the request to the Data Source if the permission state of the requested data indicates an active permission. As noted in the following chapter, the MyData Operator is responsible for this introspection functionality.
- Compliance requirements set demands for the Data Source (as the original controller) to prove audit trail of the personal data use over the data it provides out. To meet these requirements, the connector needs functionality to store a local audit trail of all data access.
- In a multi-operator environment, the audit trail SHALL be provided so it covers all MyData Operators the connector is and has been using for identity and permission management functionalities.
- In a multi-operator environment, the audit trails SHALL be separable on a per-operator basis to allow statistics for e.g., measured charging of data access on a per-operator, per Data Using Service basis.

- The Data Source makes available to the trust group, through the connector, its data sharing policies that can be matched with any Data Using Service's profile connected to the trusted group. *(requirement for future scope with machine readable agreements)*
- The Data Source makes available, through the connector, the data sharing agreements that it already has for operators to check compliance of data transfers and enable them. *(requirement for future scope with machine readable agreements)*

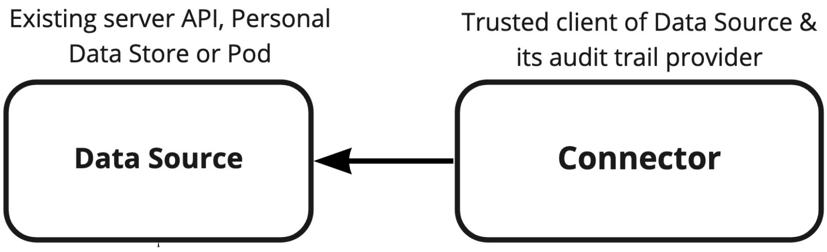

Existing server API, Personal Data Store or Pod
Trusted client of Data Source & its audit trail provider
Data Source
Connector

Figure 20: Relation between Data Source and a MyData connector

An alternative approach to a proxy client implementation is to implement the required connectivity functions (see Section 5) within the Data Source itself. In this case the relation diminishes what could be illustrated conceptually as in Figure 21.

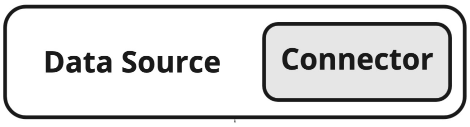


Figure 21: Data Source with built-in connector support

Note

Figure of an integrated data source and connector is used also in figures of this specification where it doesn't make sense to keep the two components separate or the implementation approach doesn't matter in the context.

4.2.2 Connector – MyData Operator relation

Like the case with Data Source integrations with a new client, the permission management services of MyData Operators expect the Data Source or its proxy to communicate through their specific permission management related protocol. This is needed when connector validates any incoming data requests from the related MyData Operator. The means to identify the MyData Operator that is handling permissions for a particular Data Using Service is described in detail in Section 5.3.2.

Specific functional requirements of the Connector – Operator relation are:

- Connector needs to be able to serve as a trusted Data Source client towards multiple different MyData Operators simultaneously. This is needed for permission validation mechanism that is based on introspecting any claimed permission from the remote operator (see Figure 22).
- Connector SHALL introspect an inbound request from a Data Using Service using the permission management functionality of the MyData Operator who the Data Using Service has used to request the permission from the data subject. The MyData Operator SHALL provide a response that indicates whether the Connector is allowed to perform the request to the Data Source.
- The MyData Operator SHALL maintain an access log of the introspection interactions and provide access to the audit trail through an API.
- Connector needs to be able to separate the permission management handling and audit trail collection for each connected MyData Operator. This collection may be organized and implemented in various ways, the detailed functionality in Section 5.6 of this specification describes a generalised solution. It is based on current audit trail generation approach of MyDataShare Access Gateway, an open-source MIM4 connector implementation described in Section 6.
- In multi-operator & multi-vendor environment flexible configurability and plug-in type of architecture of the connector is needed for support of multiple permission management approaches.

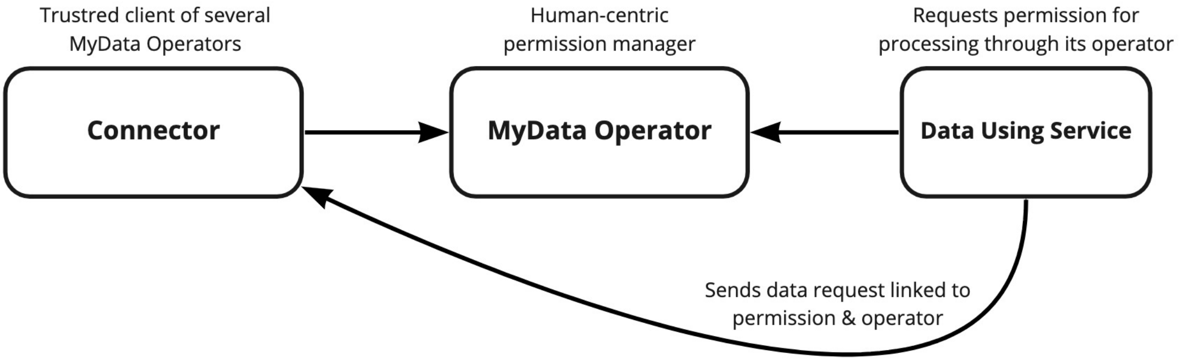

Trustred client of several MyData Operators
Human-centric permission manager
Requests permission for processing through its operator
Connector
MyData Operator
Data Using Service
Sends data request linked to permission & operator

Figure 22: Connector facing a single data request and related permission validation with a MyData Operator.

4.2.3 Connector – Data Using Service relation

A connector represents a Data Source towards MyData services that want access to data it provides. Technically this means that instead of protecting the actual Data Source against 3rd party access requests via an existing access management system, any incoming MyData-related traffic is routed via the connector.

Specific functional requirements of Connector - Data Using Service relation are:

- Connector itself MUST be a trusted client of the Data Source, if not implemented by the Data Source itself

- Connector itself MUST be accessible by the Data Using Services via an *endpoint* which is published (made discoverable) to ecosystem parties
- An accessible endpoint or route may be specific to a MyData Operator. That is, there may be several routes served by the Connector ingress API. Each route may be dedicated to data request traffic related to a specific MyData Operator.
- to assist the developers of Data Using Services, the format for any incoming requests to a particular API endpoint MUST be documented in public domain, e.g., via OpenAPI specification referred to by a publicly accessible URI.

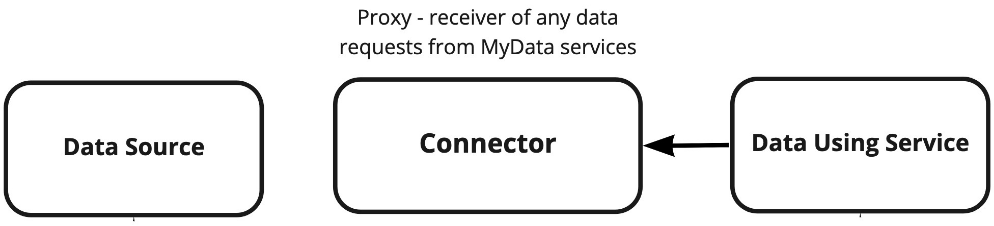


Figure 23: Relation between connector and a data using service that belongs to the ecosystem.

4.2.4 Trust Group Registry – Connector relation

A need for the connector to *introspect whether an incoming call arrives from within a trust group which the Data Source itself is part of* is functionality that the Personal Data Management flows within a single-operator ecosystem typically don't cover. All operational factors of trust group registry (who runs and maintains it) are not part of this specification.

Section 5.1 specifies a simple web-based solution for a trust group registry function. The underlying functional requirements between the components are:

- [an independent party/MyData Operators of the group jointly] SHALL maintain a shared list of MyData Operators that belong to an identified Trust Group.
- MyData Operator as a member of a Trust Group SHALL provide the trust group maintaining entity a link to its public key. The list of operators (operators' identifiers and their public key URLs) is published via the trust group registry.
- [an independent party/MyData Operators of the group jointly] SHALL provide trusted access to the trust group registry to shared connectors. This is required to realise the necessary read-only trust group member query mechanism specified in Section 5.3.2.
- MyData Operator as a member of a trust group SHALL maintain a shared list of the Data Sources (Connectors) which it has agreed to share within the trust group. This shall be tied to specific data sharing agreements that can be traced.
- The list of Data Sources is published via a publicly accessible mechanism. Note: Any operator may also have private (non-shared, non-published) Connectors and Data Sources.
- Connector SHALL be configured with an API endpoint of each Trust Group it belongs to.
- Connector SHALL validate via local data caching and when necessitated by cache expiration via the above endpoint(s) that any incoming data request from a Data Using Service is coming from context of an MyData Operator that belongs to at least one of the connector's configured trust groups.

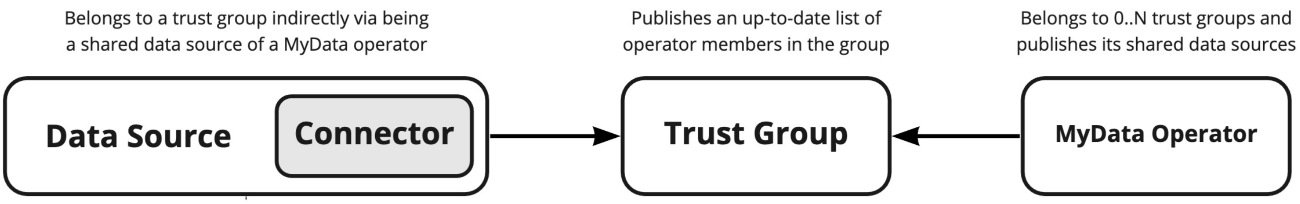

Belongs to a trust group indirectly via being a shared data source of a MyData operator
Publishes an up-to-date list of operator members in the group
Belongs to 0..N trust groups and publishes its shared data sources
Data Source
Connector
Trust Group
MyData Operator

Figure 24: Functional relations between Data Source, operator and a trust group

First version of this functionality may provide a simple pull mechanism for the Connectors to use. At a later stage MIM4 needs to create a more sophisticated mechanism for the registry to inform the members of a trust group dynamically of changes in its included MyData Operators.

5 Technical specification and sequence diagrams

This section specifies in component-level the features required to realise the requirements and functionalities listed in Section 4 above.

5.1 Trust Group registry

Trust Group in context of MIM4 connectivity is an online registry that can be maintained by any entity the trust group members (MyData Operators) together agree to trust as a neutral registry provider.

Registry can be implemented as a web service with an API endpoint that serves out a trust list object representing the group of currently valid Trust Group members. Registry provider (entity with administrative write/patch access to the registry) SHALL keep the registry's trust list updated with agreed frequency. Clients (Connectors) reading data from the registry SHALL follow any caching guidelines agreed between the group members.

Note

A single trust group's contents should not change too frequently - only when new members join, or old members are leaving a particular operator group. The Connector can therefore safely assume caching of fetched trust lists is accepted for **24 hour**s, i.e., check the list contents from the registry only once per day per group.

5.1.1 Operator description

A *trust group* consists of two or more operatorDescription arrays bound a trust group identifier.

Each operatorDescription is a JSON object which SHALL include the following contents: operator identifier *operator_uuid* and operator base URL *operator_base_url.*

- The operator identifier SHOULD be unique among MyData Operators, thus the vendors and operators MUST create their identifiers using version 4 UUIDs.

- The operator base URL links to the operator’s public metadata (see RFC 8615) at the Internet and helps dependent parties to resolve required operator endpoints and other parameters.

Attribute	**Type**	**Notes**
`operator_uuid`	string	Operator’s unique identifier
`name`	string	Operator name
`operator_base_url`	string	Operator’s base URL that serves as the public resource for operator’s endpoints and parameters. Base URL SHALL point to the root which hosts the operator’s `.well-known/mydataoperator-config` information that contains its MyData Operator specific metadata. **Note:** *Formal registration of the .well-known URI with IETF to be conducted by OASC.*

An example of a single operatorDescription as JSON object:

```
{
        "operatorDescription": {
        "operator_uuid": "5e00c14239024fe1b8175713ae1f77cb",
        "name": "Acme",
        "operator_base_url": "operator.acme.com/"
  }
}
```

Note: The elements contained in operatorDescription doesn’t need to be served by a singular operator locally (not hosted in a Trust Group registry) to inform outside world about its unique operator identifier and public configuration metadata endpoint. The two elements are available behind the operator’s .well-known URI (see mydataoperator-config in Section 5.7).

However, the MyData Operator identifier of a list member is important to store conveniently within the registry, within a single request for the Connector to avoid unnecessary network traffic prior the Connector can make its operator_uuid attribute value -based decision on trustworthiness of a standing data request.

5.1.2 Groups endpoint

Trust group registry SHALL provide the trust list object at a public API endpoint served at resource `[https://example.com]/trustlist-api/groups`. Read-only scope (GET method) to the endpoint SHALL be provided to all others than the assigned registry provider.

5.1.3 Trust group object

A *trust group* consists of two or more `operatorDescription` objects grouped as `members` of a group. To identify a Trust Group, a `trust_group_identifier` is defined as a Version 4 UUID. It SHALL be created by the registry provider upon setting up the group.

Attribute	**Type**	**Notes**
`Trust_group_uuid`	string	Trust Group's unique identifier

Trust group members are available as JSON object via the Trust Group registry's public endpoint (see groups endpoint).

Trust group object payload SHALL be encoded and signed as JSON Web signature (JWS, see RFC7515) using the trust group registry provider's private key, and can be decrypted by any holder of the trust group registry's public key. JWS JSON Serialisation SHALL be used.

Example of a decoded Trust Group with two MyData Operator members:

Method and parameters	**Response (decrypted)**

GET
example.com/trustlist-api/groups
	{
	"trust_group": {
	"trust_group_uuid": "07193772-f433-43d4-83bf-b34fcc6ac8e1",
	"members": [
	{"operatorDescription": {
	"operator_uuid":"f240fcf4-d0bb-4b3a-8779-e7099e68d104",
	"name": "Example1",
	"operator_base_url": "operator1.example.com"
	}
	},
	{"operatorDescription": {
	"operator_uuid":"dd56957e-bf80-4dbd-ac5d-e0f4c7d5187e",
	"name": "Example2",
	"operator_base_url": "operator2.example.com"
	}
	}

	] } }

5.2 Data request process

The primary use case for a Connector is to shield a Data Source by verifying that an inbound data request is valid. The Connector accomplishes this by introspecting the data request using the permission management functionality of the MyData Operator who is responsible for the Data Using Service.

As a preceding process to this, the Data Using Service creates a permission request using an operator-specific API. A permission request ties together the Data Using Service, a Data Source and the Individual whose data is targeted. The Data Using Service shall utilise the identifier of the permission request to perform the actual data requests to the Data Source.

The following figure outlines the sequence of a data request to a Data Source which is shielded by a Connector.

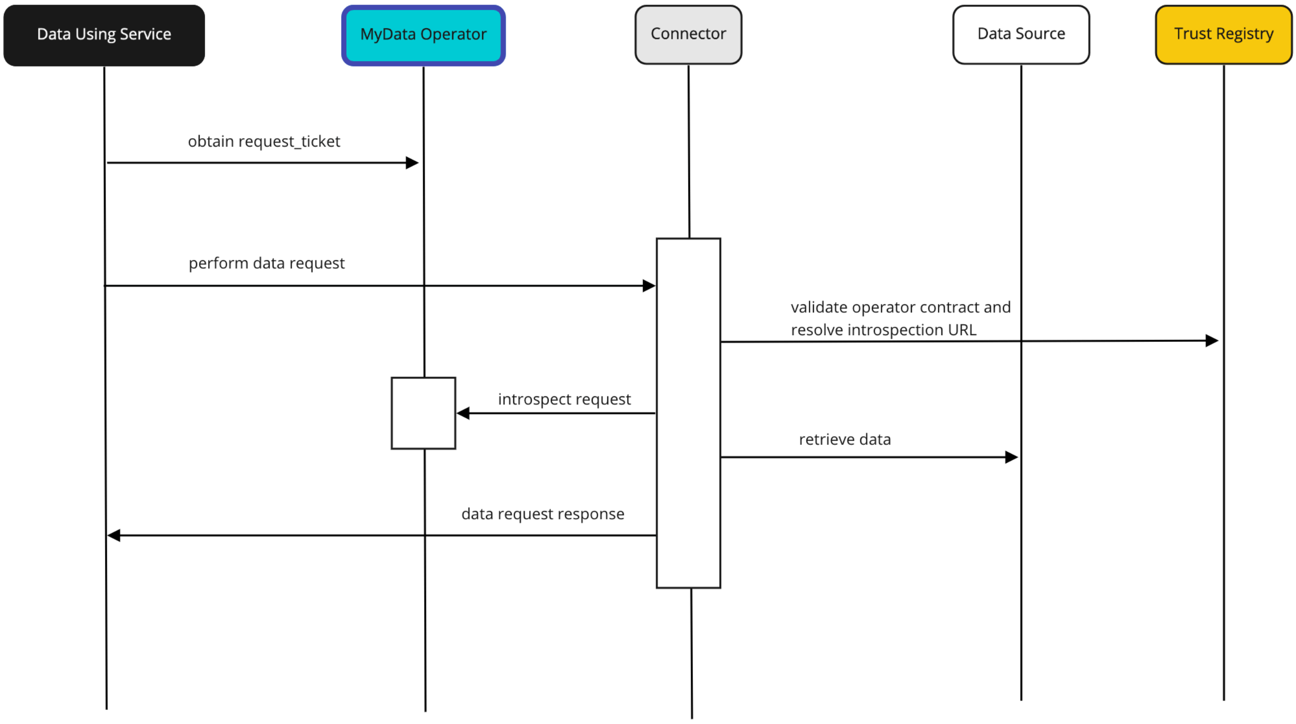


Figure 25: Connector-shielded data request to a Data Source

1. The Data Using Service obtains a request ticket from the MyData Operator[24].
2. The Data Using Service formulates the data request. Instead of sending the request directly to the Data Source, it sends the request containing the request ticket to the Connector instead.
3. The Connector resolves which MyData Operator's permission management service needs to be utilized to validate the request by utilizing the Trust Registry service or its own contracts (as described in 5.3.2), and routes the request ticket to the appropriate operator for validation.
4. The MyData Operator utilizes the request ticket payload to verify that the Data Using Service is allowed to perform the request to the Data Source - essentially this is done by checking the state of the permission request. If the Individual has granted permission to the Data Using Service, the introspection response indicates this. The MyData Operator stores the introspection result in its audit trail.
5. Upon receiving a positive response, the Connector performs the request to Data Source, receives the response from the Data Source and finally formulates the response and sends it to the Data Using Service.

Additionally, to simplify definition of request-management, the Connector may expose operator-specific endpoints instead of resolving the used MyData Operator on a per-request -basis.

5.3 Request ticket

A request ticket is a token which allows a Data Using Service to perform data requests to a Connector-shielded Data Source for the personal data of an Individual.

A request ticket is realized as a signed JSON Web Token (JWT), which consists of the following claims:

Claim	Type	Notes
`iss`	registered	Issuer = operator_uuid, the Operator identifier
`sub`	registered	Subject = Organization that owns Data Using Service
`aud`	registered	Audience = URL of the Connector endpoint targeted
`exp`	registered	Expiration = UTC unixtimestamp
`iat`	registered	Issued at = UTC unixtimestamp
`jti`	registered	JWT id = unique identity of the ticket
`mop_processing_record_uuid`	private	Example of an operator-specific payload

[24] It is for future scope of this specification to find approach to validate if the Data Source and Data Using Service have a data sharing agreement before the MyData Operator before provisioning a request ticket to Data Using Service.

Storing version identifiers for both the registered and operator-specific claims could be a useful extension. However, as a JWT parser should be implemented in such a fashion (iterate over the contents of the token) to mitigate this need, this API has been left unspecified.

If version identifiers are seen as an idea to pursue further, the version identifier for the registered claims should be registered as a public claim, with OASC as the registering organization (see RFC 7519, chapters 4.2 and 10.1 for details).

5.3.1 Obtaining a request ticket

A Data Using Service obtains a Request Ticket from a MyData Operator by performing an operator-specific request and supplying the required parameters as listed above. Each operator vendor implementing this specification SHALL document and release their own required private parameters for use by Connectors that want to support multiple operator vendors/types.

Note

In case of Vastuu Group's MyDataShare the required parameters correspond to the contents of the operator-specific content of the request ticket: the identifier of the permission request. See vendor-specific API reference documentation: https://app.swaggerhub.com/apis/MyDataShare2/MyDataShare/.

Add other vendor-specific notes here, when known.

5.3.2 Validating trust status of request ticket

Upon receiving a request ticket, the Connector SHALL verify that the identified operator is either a member of a Trust Group with which the Data Source has a valid contract, or a MyData Operator with which the Data Source has a valid contract outside any group context.

5.3.2.1 Trust group query mechanism

Trust group management is based on the Trust Group members deciding where in the Internet they are to host the Trust Group registry function. Each MyData Operator member of a group SHALL provide this Trust Group registry URI and its public key to their shared Data Sources as part of their Connector configuration information.

To verify status of the identified operator, the Connector SHALL send a request to the groups endpoint of a Trust Group registry URI it has been configured with, and decode the received JWS object with the public key of the trust group registry.

If the `iss` claim value of the request ticket can be found among the `operator-uuids` in the decoded trust list object, the request ticket originates from a group member and the Connector can proceed to introspect the permission with this operator.

In case the Connector doesn't find the `iss` claim value from the trust list or it lacks a Trust Group registry URI in its configuration, it SHALL compare the value with the operator identifiers it has known agreements with. This is done via the linked operator's `mydataoperator-config` metadata.

If the request ticket iss attribute value doesn’t match with any operator_uuid known or resolvable by the Connector, the data request SHALL be rejected as unauthorised.

5.3.3 Future extensions

For dealing with machine-readable data sharing agreements, the following type of private claim may be introduced to the request ticket in the future:

Claim	Type	Notes
contract_token	private	Token enabling to prove and check the existence of a data sharing agreement between Data Source, Data Using Service, MyData Operator and Person (if relevant).

5.4 Permission introspection

Upon the Connector obtaining the identity of the MyData Operator and verifying that it has a valid contract with that operator, it requests the introspection of the request ticket from the MyData Operator. The Connector SHALL resolve the URL of the operator’s introspection endpoint dynamically from the MyData Operator’s public mydataoperator-config metadata.

The request SHALL contain the request ticket, and the body of the response SHALL be a JSON dictionary and contain the following elements:

Element	Description	Notes
active	Whether the introspection allows the Connector to proceed with the request to the Data Source (if true)	
reason	A human-readable message that provides additional information in case of errors	
access_item_uuid	Identifier of the audit trail element created for the introspection.	Empty if introspection does not allow data request to be performed.
identifiers	Array of Individual’s personal identifiers that the Connector is allowed to utilise in performing the data request towards the Data Source.	

An identifier is a JSON dictionary and contains the following elements:

Element	Description	Notes
`id`	Value of the identifier	e.g. “220279-161A”
`id_type`	Type of the identifier	e.g. “ssn”, “passport_number”, “email”
`country`	Country that has issued the identifier (value present only if the identifier is issued by a national issuer)	e.g. “FIN”
`verified`	Timestamp when the identifier has been verified	

5.5 Use of personal identifiers

As noted in the previous chapter, the MyData Operator SHALL provide Connector a list of the personal identifiers that the Connector CAN utilise when performing the client request towards the Data Source.

The personal identifier used by the Data Using Service (required at the time of initiating a permission request via the MyData Operator, usually orchestrated with invoking a privacy notice for the Individual and authenticating her if this was not done earlier in the process) SHALL NOT need to match with the personal identifier linked to the Individual at the Data Source.

This approach is to minimise the possibility for identity leaks and related fraud within the Personal Data Management.

Design principle for personal identifiers

The identity management function of the MyData Operator should keep a private mapping of an Individual’s linked identifiers at different Data Sources and Data Using Services keyed to the Individual-keying identifier created at the operator itself (defined herein as ‘base identifier’).

Base identifier of an Individual at the operator (or other service specific keying identifiers) should never be a personable identifier such as GovID or social security number. Such may have to be used for authentication of an Individual at the time they make an authoritative decision regarding their personal data use, but only as an *identity attribute, not as the base identifier* of an Individual in a service or application.

5.6 Audit trail

The specified MyData ecosystem contains two separate audit trail logs that contain entries about introspection interactions and data requests:

- Operator log (maintained by MyData Operator)
- Data Source log (maintained by Connector)

These two separate logs are necessary. A MyData Operator's log is used by the operator, by the Data Using Services and the Individuals. Additionally, a Data Source may be used by multiple MyData Operators, in which case its interactions will be split into multiple MyData Operators' logs. Instead of the Data Source harvesting the logs from the operators, it is easier if it maintains a corresponding log of its own.

Additionally, the MyData Operator also maintains an audit trail log that contains entries about the lifecycle and status transitions of permission requests (and other entities, when feasible).

5.6.1 Operator log (MyData Operator)

The MyData Operator SHALL create an access log entry upon completing introspection of a request ticket to record the introspection result.

The Connector SHALL update the access log entry upon completing the request to the underlying Data Source and responding to the Data Using Service. This update is done using the `access_item_uuid` identifier provided in the introspection response.

5.6.2 Data Source log (Connector)

Minimally the Connector SHALL store the introspection response and the status of the Data Source request and response to the Data Using Service.

However, the Connector MAY store additional information on account of architectural or operational reasons.

5.7 Operator configuration metadata

Operator must provide a description on its configuration and endpoint locations. For this the operator MUST implement a `/.well-known/mydataoperator-config` endpoint. Information on this endpoint SHALL be served as a JSON object, which contains the following attributes about the Operator:

Attribute	**Type**	**Notes**
`operator_uuid`	string	Operator's unique identifier (UUID v4)
`operator_key`	JWK	Operator's public key. JWK structure MUST contain 'kid' parameter.
`name`	string	Name of the Operator
`vendor`	string	Operator vendor
`operator_base_url`	string	URL which serves as common entry to the Operator, and hosts the metadata of this object
`introspection_url`	string	URL of the Operator's introspection endpoint, which SHALL be used by Connectors for permission introspections.

		In case this is given as a relative path, the introspection endpoint is `operator_base_url` appended with `introspection_url`.
`api_guide`	string	URL to the vendor's OpenAPI documentation
`shared_connectors`	object array	JSON object array of Operator's shared Connectors. (optional)
`restricted_connectors`	object array	JSON object array of Operator's Connectors whose visibility is restricted to Data Using Services of its own. (optional)

5.8 Connector configuration metadata

A similar metadata publication for the Connectors as above for Operators is a useful mechanism to serve decentralised Connector discovery. Entity behind a shared Connector that resolves from the metadata of its primary Operator can release their up-to-date configuration information. Adoption of this approach contributes to realising some parts of the Service Management function described in the MyData Operator white paper.

A Connector must provide description on its configuration and endpoint locations. For this Connector MUST implement `/.well-known/connector-config` endpoint. Information on this endpoint SHALL be served as a JSON object, which contains the following attributes about the Connector:

Attribute	**Type**	**Notes**
`connector_uuid`	string	Unique identifier of the Connector. UUID v4 SHALL be used.
`connector_key`	JWK	Connector's public key. JWK structure MUST contain 'kid' parameter. (optional)
`name`	string	Name of Connector
`description`	string	Human-readable description of the Connector
`api_guide`	string	URL to the Data Source's public OpenAPI documentation. This documentation MUST explain how Data Source provides access to its resources via the Connector.
`connector_base_url`	string	URL which serves as common entry to the Connector itself

5.9 Shared Connectors of a MyData Operator

The Operator's metadata configuration served via the .well-known endpoint can be used by the Operator to disclose the Connectors it currently shares out to its identified Trust Groups. In such case, the `connector_base_url` of a shared Connector MUST be listed on a `shared_connectors` object array.

A Connector may belong simultaneously to multiple Trust Groups, so it MAY be listed under one or more Trust Group identifiers.

Example of a `shared_connectors` object:

```
{
        "shared_connectors": {
                "trust_group_uuid": "07193772-f433-43d4-83bf-b34fcc6ac8e1",
                "connectors": [
                        {"connector_base_url": "connector1.example.com"},
                        {"connector_base_url": "connector2.example.com"},
                        {"connector_base_url": "connector3.example.com"}
                ]
        }
}
```

5.10 Restricted Connectors - *To be specified later*

Access to resolve Operator's private Connectors may be defined with a vendor-specific method. This can be a similar list as defined in Section 5.9 for shared Connectors, but only resolvable through a non-public endpoint that the MyData Operator provides for its registered clients.

5.11 Publishing Connector resource descriptions - *To be specified later*

6 MyDataShare Access gateway as a Connector implementation

MyDataShare Access Gateway (hereafter AGW) is a reference implementation of Connector.

AGW is designed using “configuration over code”-principle. Thus, a Data Source may define the behaviour of a route using building blocks provided by the implementation as shown in the example below:

```
{
    "route": {
        "path": "/foo/simple/<baz>",
        "method": "POST",
        "plugins": [
            { "plugin": "mop_request_ticket_validation" }
        ]
    },
    "requests": [
        {
            "url": "http://dataprovider.example.com/foo?item=0&limit=100",
            "method": "POST",
            "headers": {
                "version": "bar ${route.headers.Version}",
                "omit": "lol ${route.query.except[0]}"
            },
            "json": {
                "request_list": "${route.baz}"
            }
        },
        {
            "url": \
           "http://dataprovider.example.com/bar?count=${requests[0].response.json.count}",
            "method": "GET",
            "headers": {
                "locale": "bar ${route.headers.Locale}"
            }
        }
    ],
    "response": {
        "status": 200,
        "headers": {
            "content_type": "application/json"
        },
        "json": {
            "count": "bar ${requests[0].response.json.count}",
            "allow": "${requests[0].response.headers.Access-Control-Allow-Headers}",
            "status": "resp1 ${requests[1].response.json.status}"
        }
    },
    "after_hooks": [
        { "after_hook": "patch_mop_access_item" }
    ]
}
```

The `route` defines the endpoint available for data consumers (as "inbound request").

The `requests` defines the set of provider requests to be performed by the AGW to harvest the data for the response from the actual Data Source.

The response defines how to generate the response to be sent to the data consumer after all the provider requests have been performed and their responses run through the processors.

The `after_hooks` defines the actions performed after the response has been generated and sent.

However, in case a Data Source requires additional functionalities, developing and integrating building blocks is not a major task (as proof for this extension concept, AGW includes a XML2JSON-conversion mechanism).

AGW is available in Github (see Annex B). It is a moderately sized python implementation that is deployed and run in containerized fashion. It can be utilized as-is or viewed as a reference implementation and implemented using another programming language and/or application framework.

6.1 Access Gateway hosting

The AGW can be hosted in two modes:

- A Data Source can self-host it.
- A Data Source can use an instance of an AGW that is hosted by the MyData Operator.

PART 3 – Evolution of MIM4

1 MIM4 Future roadmap and extensions

In the earlier MIM4 related documentation the concept of a MyData Operator and personal data platforms were introduced. This MIM4 version has its focus on connectivity and legal framework. The intention is that these will provide the first necessary components to progress to an implementation phase with MIM 4 and to take personal data connectivity to the next level.

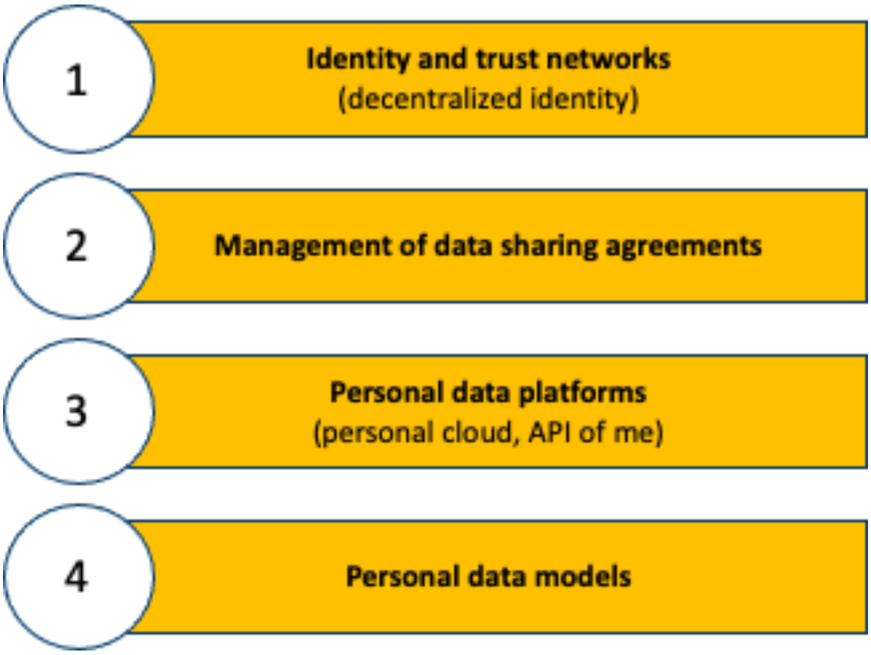

1
Identity and trust networks
(decentralized identity)
2
Management of data sharing agreements
3
Personal data platforms
(personal cloud, API of me)
4
Personal data models

Figure 26: Future roadmap includes improving specifications on Levels 2 and 3, and to introduce new aspects for Level 1 and 4.

However, this current version of MIM4 specification is only the first step towards a more comprehensive MIM specification. There are several aspects under the current focus of the specification that will need further work. Also, new areas that address identity and trust networks (Level 1) and the personal data models (Level 4), are envisioned in future MIM4 versions.

1.1 Roadmap for MIM4 connectivity specification

The requirements regarding the MIM4 connectivity specification, which are to be implemented in future versions are listed in the following table.

ID	Requirement	Actors	Source

B.I	Support multiple types of personal Data Sources beyond currently implemented centralised data registries (e.g., SOLID, decentralised Data Sources, etc.)	• Data Source	• Service Creation
B.II	Define federated permission management between multiple operators.	• MyData Operators	• Service Creation
B.III	Define federated audit trail and logging between multiple operators.	• MyData Operators	• Service Creation
B.IV	Define required federation and roaming (e.g., connectivity, id federation, permission management) between multiple Data Sources and/or MyData Operators	• MyData Operators • Data Sources	• Service Execution • Connect a new Data Source to a MyData Operator
B.V	Define federated governance principles between entities under separate rule books / governance models	• MyData services • MyData Operators • Data Sources	• Service Execution • Connect a new Data Source to a MyData Operator • Adding a new MyData Operator to a service
B.VI	Define settlement principles between Data Source and multiple MyData Operators	• MyData Operators • Data Sources	• Connect a new Data Source to a MyData Operator
B.VII	Define required federation and roaming (e.g., connectivity, id federation, permission management) between multiple MyData Operators using a primary MyData Operator.	• MyData Operators	• Adding a new MyData Operator to a service
B. VIII	Define interactions with the legal component implementing the MIM4 layer 2 for data sharing agreements	• MyData services • MyData Operators • Data Sources	• Service Creation • Service Execution

1.2 Legal component for automated contracts

The legal component implementing the machine-readable contracts is characterized by different functionalities, which can be ensured by the MyData Operator or the Data Source through its connector:

- **Define and publish data policies (DP):** for each data set the Data Source can specify conditions, obligations, restrictions, prices, certifications, data security, rights, data protection, liability; these provisions to access a data set define the data policy, are made machine readable using appropriate standards and are made available (published) to the members of the data network.

- **Generate and manage data sharing agreements (DSA):** when a Data Using Service agrees and complies with the data policy, a data sharing agreement (DSA) is generated between the Data Using Service, the Data Source and the MyData Operator taking into account the data policy's provisions. The DSA is generated in human and machine-readable format, using appropriate standards, and its metadata (parties, data set, provisions, time) are made available to the parties of the data sharing agreement.
- **Proving data sharing agreements:** a member of a data network can check the existence of a data sharing agreement, get relevant metadata about the agreement and a token proving the validity of the agreement that can be used to enable data exchanges.

This will ensure that DSAs can easily be generated, that each data transfer is tied to a DSA and that each data transfer does not happen without the validity of a DSA being proven.

Such a component is under development under the EU-NGI DAPSI Rulebook project, and first versions and specifications will be available in Q3 of 2021.

1.3 MyData Operator Network Rulebook

Requirements and roadmap for a multi-operator network are defined in MyData Operator Network Rulebook that is to be implemented in Q3/2021. This rulebook that governs the collaboration between MyData Operators provides a cross-mapping between this MIM4 Connectivity Specification and the Service Management function as defined in the MyData Operator White Paper.

REFERENCES

NAME	DESCRIPTION	REFERENCE
7 Tech ways to get started with SynchroniCity and the OASC MIM principles	Synchronicity and its relation to OASC minimum interoperability mechanism (MIM) principles.	https://synchronicity-iot.eu/7-tech-ways-to-get-started-with-synchronicity-and-the-oasc-mim-principles/
Advanced Notice & Consent Receipt – ANCR-WG	Advanced Notice and Consent Receipt (ANCR-WG) • Working to update existing V1.1 of the Consent Receipt Specification to match GDPR and other legal requirements as well as fulfill technical gaps. Started in January 2021, expected to be a six-month effort. Will be published as a Kantara Specification. • Other related actions: Notice and Consent Task force at ToiP(https://trustoverip.org/), DIACC (https://diacc.ca/), aNG, ISO, W3C DPV	https://kantarainitiative.org/confluence/pages/viewpage.action?pageId=140804260
Data Privacy Vocabulary (DPV)	Data Privacy Vocabulary (W3C DPV): • GDPR-extension: https://dpvcg.github.io/dpv-gdpr/ • GIT (includes e.g. the JSONLD definition of the DPV): https://github.com/dpvcg/dpv/	https://dpvcg.github.io/dpv/
GDPR Regulation	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)	https://eur-lex.europa.eu/eli/reg/2016/679/oj
MIM4: Personal Data Management	OASC wiki pages for Personal Data Management	https://oasc.atlassian.net/wiki/spaces/OASCMIM/pages/30179329/MIM+4+Personal+Data+Management
MyData Architecture – The Stack	MyData architecture and technical specifications	https://hiit.github.io/mydata-stack/
MyData Declaration	MyData declaration and related materials	https://mydata.org/declaration/
Open API Table	TM Forum Open APIs and component suites provide service and a technology-neutral suite of APIs that provide the minimum building blocks for interoperability across all operational management areas. Each API and component suite provide the specification, reference implementations and in most cases conformance test kits. Reference Implementations are available under the Apache2.0 license. These APIs have gained global adoption in the Telecommunications industry and are proven to maximize reuse. They are designed to be extendable as required for	https://projects.tmforum.org/wiki/display/API/Open+API+Table

	specific services. The respective data models have been harmonised with FIWARE and GSMA data models.	
Rulebook for a fair data economy	IHAN fair data -based rulebook	https://www.sitra.fi/en/publications/rulebook-for-a-fair-data-economy/
Solid protocol	Solid Specifications for providing applications with secure and permissioned access to externally stored data in an interoperable way.	https://solidproject.org/TR/protocol
Streamlining governmental processes by putting citizens in control of their personal data	SOLID backgrounder paper	https://ruben.verborgh.org/publications/buyle_egose_2019/
SynchroniCity: Delivering an IoT enabled Digital Single Market for Europe and Beyond	Basic Data Marketplace Enablers by Synchronicity.	https://synchronicity-iot.eu/wp-content/uploads/2018/09/SynchroniCity_D2.4.pdf
Testbed for fair data economy	IHAN as testbed for fair Data economy, introduction	https://www.sitra.fi/en/projects/testbed-for-fair-data-economy-ihanfi/
The City of Helsinki Data Strategy	Data strategy for the City of Helsinki	https://digi.hel.fi/english/helsinki-city-data-strategy/
Understanding MyData Operators	An introductory paper to MyData Operators	https://mydata.org/wp-content/uploads/sites/5/2020/04/Understanding-Mydata-Operators-pages.pdf

GLOSSARY

This glossary explains some of the key concepts used in the MIM4 specification documentation.

TERM	DESCRIPTION
Actor	An organisation or an individual performing one or more roles.
Data Consumer	The role responsible for processing personal data from one or more Data Sources to deliver a service. In this document, Data Consumer is synonymous with the term **Data Using Service.**
Data Ecosystem	A system of interrelated Data Networks.

Data Governance	A system that employs interoperability components (standards and policies) to ensure the acceptable use and high quality of data within a specific ecosystem. Manages the availability, usability, consistency, integrity, and security of the data used.
Data Intermediary	A provider of data sharing services as defined in the Regulation on European data governance (Data Governance Act) or its current draft.
Data Network	A group of organizations and/or individuals that build data sharing solutions. / A group of companies and other organizations or even individuals that share data in accordance with a rulebook or other contractual arrangement.
Data Portability	The ability of data to be easily moved across interoperable applications and domains. The legal right to data portability, granted in some jurisdictions to individuals, can be delivered through a range of technical mechanisms and varies in scope according to the jurisdiction. The MyData principle of data portability encompasses the ease of both access to and reuse of data.
Data Provider	Any natural person or an organisation that provides data for the parties to use via a Data Network. Note that in MyData specifications this is synonymous to Data Source.
Data Source	Any source system for data. For example, weather APIs (application programming interfaces), internal systems and databases, IoT devices. Note that in MyData specifications this is synonymous to Data Provider.
Data Using Service	The role responsible for processing personal data from one or more data sources to deliver a service.
Fair Data Economy	A fair data economy accommodates the interests of all types of participants while also providing for a high level of overall data usage. In a fair data economy: ● Individuals know how their data is being used, can freely give and revoke required permission for the use of their data and mandate its sharing with third parties. They gain a share of the benefit from their data, typically not in monetary form but in terms of better services. ● Service providers include for instance social media, banks, utilities, hospitals and retailers. They share control of their users' data, often investing significant resources to co-produce them. They are able to share personal data with third parties based on a range of legally valid reasons, including consent. They may need to provide their customers with portability rights, but also be able to build innovative services on users' data. A fair data economy is not a form of data collectivisation: it does not require service providers to give up and share their aggregate Data Products as such, only individual data through portability. ● Data re-users are able to access a customer's personal data hosted by the service provider to provide them or others with new services. Data should not constitute an excessive barrier to entry. And researchers and innovators should be able to make the best out of the data. Data re-users include for instance third-party payment providers or independent businesses that directly compete with the service provider, but also other parties, such as data analytics companies or researchers, that are in different lines of business and can innovate by re-using the data. Both service providers and data re-users are accountable for misusing personal data.
GDPR	General Data Protection Regulation, i.e. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Operator	The role responsible for operating infrastructure and providing tools for the person in a human-centric system of personal data exchange. Operators enable people securely to access, manage, and use personal data about themselves as well as to control the flow of personal data within and between data sources and data using services.
Person	The role of data subject as represented digitally in the ecosystem. Persons manage the use of personal data about themselves, for their own purposes, and maintain relationships with other persons, services, or organisations. The words Individual and Citizen are also used in this document as a synonym for a Person.

Sources:

- Sitra Fair Data Economy Rulebook (v.1.2)
- MyData Operator White Paper
- International Data Spaces (IDS) RAM 3.0
- EU Digital Governance Act (DGA)
- Mapping of terms used in various initiatives can be found at https://tinyurl.com/jbk89bah

ANNEX A – example use case from smart city context

To be added later

ANNEX B - MIM4 connectivity baseline reference implementation

A reference implementation based on the current MIM4 specification has been developed as part of the specification effort. It is available separately from the authors and is released under open-source software license terms.

Once the implementation is available, the code repository at https://github.com/MyDataShare/access-gateway will be set available for the public.