

Discussion Paper

What is the MyData Operator?

Contributions by Antti Poikola, Joss Langford, Mika Huhtamäki, Mikko Sierla, Wil Janssen¹

Goal of this discussion paper

In the MyData declaration the Personal Data Operator has a prominent position. Still, its exact role in the eco-system is not well delineated. What are typical functions it should fulfil? What responsibilities should it have? What are valid business models for operators, for example, may it profit from data? And what do the MyData principles imply for an operator? In this discussion paper we want to coin a number of important design issues for operators. It serves as a starting point for a workshop on the role of the MyData operator at the 2019 MyData Conference. From there, we want to create a white paper to create a joint understanding of the MyData Operator in all its appearances, in collaborative writing effort. We strongly invite people to contribute to the discussion and the white paper.

What is the operator and why is it a good idea?

As the importance of personal data in society continues to expand, it becomes increasingly urgent to make sure individuals are in a position to know and control their personal data, but also to gain personal knowledge from them and to claim their share of their benefits. [MyData declaration] This is quite different from the early days of the internet, where internet pioneers with all the services you'd need, including a digital identity, even for free! The only thing they wanted in exchange for all those marvelous services was your data... Such bundling of services, however, inhibits competition, reduces the power of the market and inhibits innovation. Without the ability to choose different services, service providers or data relationships, there is no meaningful consent and the individual has no power to influence the market.

This is where the personal data operator comes in: the operator plays the role of an intermediary between data sources and data using services within a trust framework or ecosystem with fiduciary responsibility for the individual. It should create trust and open up a user driven market where individuals are both protected and empowered to use the data that organisations hold about them.

A data source collects and processes personal data which the other roles (including persons) may wish to access or use. A data using service or data consumer can be authorised to fetch and use personal data from one or more data sources. A personal data operator enables individuals to securely access, manage and use their personal data, as well as to control the flow of personal

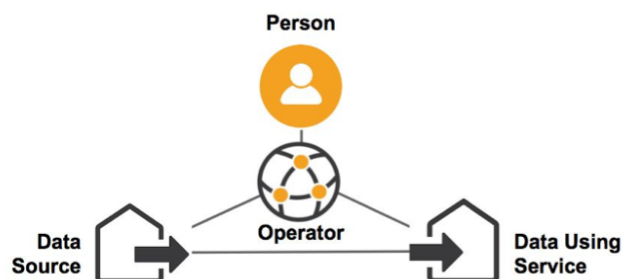


Figure 1. MyData roles in the eco-system.

¹ Contact: antti.poikola@teknologiateollisuus.fi, joss@coelition.org, mika.huhtamaki@tilaajavastuu.fi, mikko.sierla@tilaajavastuu.fi, wil.janssen@innovalor.nl

data with, and between, data sources and data consumers. The operator puts the person on the drivers' seat doing the decisions (managing) on what his/her data should be used for. A person can partly offload some of this management to the operator, so that operator can make automatic decisions on behalf of the person, based on some preferences (i.e. I will never grant permission to sell my data for third parties).

Operators come in many forms and under many different names: personal data services, personal information management services (PIMS), fiduciaries, data banks. Sometimes they manage personal data, as a kind of vault, in other cases they primarily facilitate the flow of data.



Figure 2. Examples of operators throughout the world

This diversity is a logical consequence of the early stage of evolution of the MyData eco-system, but also hampers adoption and growth. The development of these solutions started more than a decade ago; MyDex dates back to 2007, as does the trust framework Qiy. A lot has happened since then; large adoption, however, has yet to come.

The operator in context

As mentioned, the operator should create trust and open up a user driven market. It lives in an eco-system of data sources and data using services, of public organisations as well as private companies. Such an eco-system can only function properly if some form of regulation or legislation or social norms exist.

In Europe, the GDPR is a solid basis for enabling data exchange and protecting privacy. In Japan, data banks are rooted in new legislation. This type of legislation is a necessity for creation of trust, but often it is not sufficient. In order to create a level playing field in the market, rules of engagement

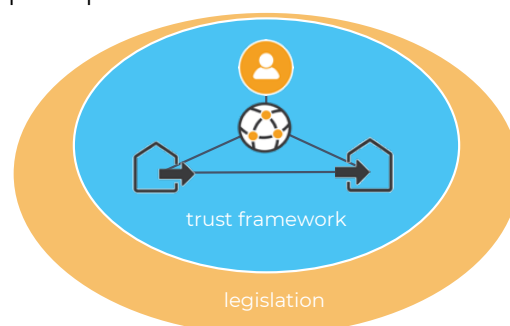


Figure 3. Eco-system builds upon trust framework, rooted in legislation.

between the different roles and actors are needed. This is often captured in a trust framework. Trust frameworks contain the underlying agreements and rule books that make a mature eco-system possible. They describe a contractually enforceable set of specifications, rules, agreements, and minimum technical specifications that govern the eco-system. They can provide regulations on data standardisation, validation of sources, consent and the portability of data. Well-known examples beyond personal data management include credit card systems (such as Visa), domain name registration systems (governed by ICANN), or telecommunication frameworks such as GSM (governed by GSMA). In this context, MedMij (Netherlands), Findy (Finland) and HAT (UK) are emerging examples of validated trust frameworks.

Often, trust frameworks are implicit or opaque to people using the services. There can be underlying contractual agreements between all parties in the eco-system, without the user needing to be directly involved. It is an interesting point of debate whether the trust framework in the context of MyData should be explicit and transparent to all parties involved, and whether or not specific restriction should be posed upon them.

A reference model

In this complex landscape, a basic, common understanding of the type of functionalities offered by operators is needed, aiding in moving from a fragmented landscape of solutions to a sustainable eco-system. For this purpose, defining a reference model is an important step. The figure below is a summary of how some of the core elements relate

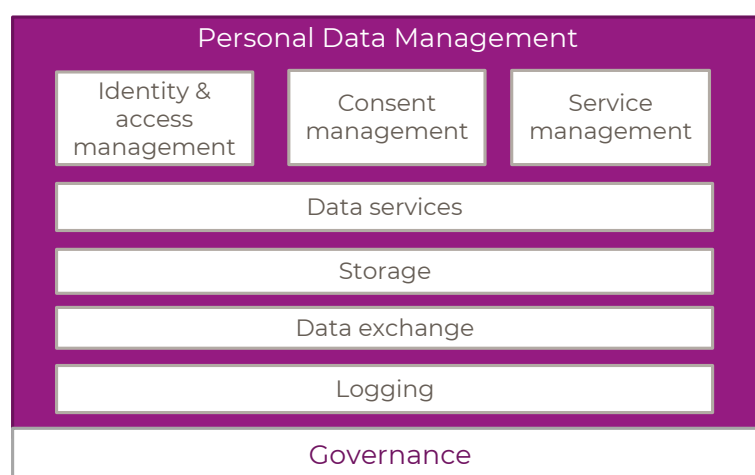


Figure 4. Personal Data Management model – top level.

to each other. The building blocks in the model have been derived from existing solutions and frameworks, such as MyData, MedMij, Qiy, digi.me and Mydex.

The operator makes it possible for a person to control which data using services receive which data. Functionality can be distributed or even duplicated over the different roles in the system: not everything resides at the operator, and some functions might apply to all roles (e.g. logging).

Important functional components to consider are:

- Consent management – managing (temporary) consents for sharing specific personal data between data sources and data consumers. Includes an overview of consent and the possibility to revoke and adapt consents given.
- Data services - the operator can add value to data by filtering, anonymizing, analysing or aggregating data, or translating data from one metamodel to another. Might also include billing for (the use of) data.
- Identity & access management - comprising authentication and authorization.
- Service management - linking data sources and data using services. Data can be available at different sources and can be used by multiple data consumers. Moreover, it can be exchanged in different ways. In a multi operator environment, it is a significant decision whether the operators use a shared service registry (potentially still distributed) or if each operator manage services separately.

- Data exchange - interfaces to allow for data exchange between data source, data consumer and operator in a standardized and secure manner. This can take different forms: structured data, supporting automated transactions or unstructured data, such as a pdf. Information can be source data, as well as derived attributes, it can be end-to-end encrypted between the data source and data consumer or processed by the data operator.
- Logging- keeping track of all information exchanges taking place, creating transparency in who accessed what and when.
- Storage - storing personal data, temporarily for the exchange or caching it. Also applies to storing user-asserted data.
- Governance – governing the use and the development and underlying principles of the ecosystem, including business model management.
- Information & transparency – providing access to and information about actors involved in ecosystems and management of an individual's data.
- Value exchange – supporting payment for services and distributing value from sharing data.

Multiple data consumers / service providers and multiple data sources will participate in the ecosystem. It is unlikely as well as undesirable that there will be 'one operator to rule them all', or even only one trust framework. Users need to be free to choose which personal data management solution is the best fit for them and the challenges they want to tackle. Data can be at many different, possibly overlapping, resources, mediated through different solutions and delivered to many different service providers. Definitely a multi-sided market. This situation will put pressure on interoperability and can be a barrier to adoption. As such, standardisation is necessary on the technical interfaces of the building blocks (APIs), and possibly even the user interface concepts. Trust frameworks such as MedMij specify technical interfaces in detail but refrain from giving guidelines on user interaction. Standards closely aligned areas such as smart cities (e.g. JTC 1/SC 27) and the internet of things (e.g. OASIS COEL) are beginning to specify the role of operator within their ecosystems.

Furthermore, there might be differences between the data exchange methods used. On the one hand, the operator might take care of the information exchange, possibly caching data, whereas for other solutions the data using services have to exchange data with the data source themselves. In many instances, the operator will hold meta-data about the individual uniquely or in parallel with other data sources.

Operator design decisions and controversies

It is of limited value to discuss building blocks and reference models of operators in an abstract way. As a way forward, we would like to focus on a number of design decision that touch upon technical, legal, business and ethical perspectives. The discussion should clarify how operators can and should adhere to the MyData principles: human centric, usability of data and open business environment. The implications of the answers to these questions should be clear to all roles involved: not only the operator, but also data sources and data using services. And, obviously, the personal perspective as well.

What should the ecosystem expect from an operator?

The ecosystem is made up of individuals, data sources, data using services and other operators. Beyond the functional aspects, what support, value and behaviour should they expect from an operator? Is independence a benefit? What are the conflicts between the needs of the individual, as presented by the operator and other actors/roles?

What determines trustworthiness?

Trust is key to operators, as they are relatively new in the eco-system and are closest to the person. What are ingredients of trust? Does a MyData operator need a trust framework? Or is transparency key? But if transparency is, how to create it in the bewildering amount of personal data and data exchange in practice?

Is the platform the operator or is there a network of operators?

Currently, some of the incumbents can be viewed as operators. They are platforms, combining many services to the users. Is this an acceptable situation if such an operator would adhere to all MyData principles? Or is a network of (interoperable) operators a necessity? If so, what would allow such a network to start, grow and sustain?



What are valid business models for operators?

No eco-system can be sustainable without a solid business model. Some value exchange needs to take place to make it happen. This doesn't imply operators should be commercial organisations, but they can be; they can still be run as a co-operation or NGO. Under what conditions, however, is it right to make money? Can data be sold or aggregated? And what is the role of the government in the business model? Should they participate financially, and in what stage of evolution of the eco-system?



What are the needs for interoperability?

If we thrive for a network of operators, interoperability becomes important. People should not be locked into a certain operator but should be able to choose and move if and when they want. What then is needed to create interoperability? Is data portability, for example using a Blue Button concept, enough, or should interoperability be promoted at technical levels, semantics or even business models? Should interoperability be the essence of the trust framework?

Are operators different in different domains, or should they be different?

Research shows that operators are adopted easier if they address a clear need. If you are ill, you may benefit significantly from a personal health environment. If your budget is under pressure, financial services might come in handy. Do operators in different domains function differently? Or can they be viewed as instances of the same generic concept? Also: should we further domain specific operators to create adoption, and thereafter work towards universal, interoperable solution?



What would make a MyData operator distinctive and valuable?

What are the important elements that a MyData operator would bring to the world? Who do we want to attract to take on the role of MyData operator? Why should organisations adopt the MyData principles and practices into their existing offerings?

What would a world without operators look like?

Operators are not the goal in themselves, they serve a role in the creation of a sustainable MyData eco-system. If operators would not exist, or if we would lack good business models for operators, how could we approach the same issues in the eco-system? What other actors can facilitate in creating a world that follows the MyData principles?



Call for contributions

This discussion paper does not provide answers, it provides a starting point for bringing together different views on the important operator concept. We would like to encourage you to contribute with your ideas and your questions, both during MyData 2019 as well as thereafter. By the end of this year we will bring together views and will try to collaboratively pinpoint different instances of operator concepts, fitting in the context they are meant for, bringing the MyData principles to life. Please contact us and join the dialogue.