Edited by:
Joss Langford
Antti 'Jogi' Poikola
Wil Janssen
Viivi Lähteenoja
Marlies Rikken

Understanding MyData Operators

Thank you

MyData Global (mydata.org) represents hundreds of individuals and dozens of organisations as members in more than 50 countries. We want to thank the members that have, through their membership contributions, made it possible to produce and publish this paper.

The following MyData Global members have supported the editorial work and production of this paper:

Coelition
InnoValor
Sitra

All MyData operators are members of MyData Global and we're grateful for their support. Awarded operators are mentioned individually at page 15 and in the Appendix 1.

A full list of MyData Global organisational members and the opportunity join as a member is available at: https://mydata.org/organisation-members

Contributors and the MyData Operators Thematic Group

This paper is a work product of the *MyData Operators Thematic Group*, a part of the MyData Global organisation. MyData Global is a registered association whose mission is to advocate for a human-centric approach to personal data.

The purpose of the *MyData Operators Thematic Group* is to develop the definition and processes associated with the MyData Operator, as described in the MyData declaration (2017, see mydata.org/declaration). The group gathers individuals and organisations with deep experience in interoperability and human-centric management of personal data. The following people have actively contributed to the development of this paper:

Benjamin André (Cozy Cloud), **Henrik Biering** (Peercraft), **Davide Calvi** (MyData Global), **Lal Chandran** (iGrant.io), **J Cromack** (MyLife Digital), **Matthias De Bièvre** (Visions), **Dominik Deimel** (comuny), **Koen de Jong** (Innovalor), **Olivier Dion** (Onecub), **Thorsten Dittmar** (polypoly), **Katryna Dow** (Meeco), **Juan V. Durá** (IBV), **Johannes Ernst** (Indie Computing), **Christoph Fabianek** (OwnYourData), **Adrian Gropper** (HIE of One), **Claudia Grytz** (esatus AG), **Bo Harald** (MyData Global), **Iain Henderson** (JLINC Labs), **Marie-José Hoefmans** (Schluss), **Jonathan Holtby** (Dataswift), **Harri Honko** (Vastuu Group), **Mika Huhtamäki** (Vastuu Group), **Wil Janssen** (InnoValor), **Kai Kuikkaniemi** (S Group), **Christian Kunz** (BitsaboutMe), **Vladimir Kuparinen** (SmartPaper.fi), **Viivi Lähteenoja** (MyData Global), **Joss Langford** (Coelition), **Jan Lindquist** (Linaltec), **Xavier Lefevre** (Fair&Smart), **Jan Leindals** (Diabetes Services), **Mark Lizar** (OpenConsent), **Lotta Lundin** (iGrant.io), **Shiv Malik** (Pool), **Alan Mitchell** (Mydex CIC), **Meindert Osinga** (Geens), **Shauna Overgaard** (Clarity Applied Intelligence), **Antti 'Jogi' Poikola** (MyData Global), **Julian Ranger** (digi.me), **Gaston Remmers** (Mijn Data Onze Gezondheid), **Marlies Rikken** (InnoValor), **Mikael Rinnetmäki** (Sensotrend), **Teemu Ropponen** (MyData Global), **Sille Sepp** (MyData Global), **Mikko Sierla** (Vastuu Group), **Dixon Siu** (Personium), **Freyja Van Den Boom** (MyData Global), **Jan Vereecken** (Meeco), **Lieve Vereycken** (MyData Global), **Maurice Verheesen** (Schluss), **Paul Wang** (ICON), **Kaliya Young** (Identity Woman), **Hadrian Zbarcea** (apifocal), **Isabelle de Zegher** (b!loba).

As the MyData Operators Thematic Group, we seek to promote the MyData operator approach to human-centric personal data management and contribute to a common understanding of that approach both within the MyData community and more widely. We bring together the best minds to provide thought leadership to inform technological and business initiatives. We focus on practical aspects of technology and governance to make the operation of infrastructures for personal data sharing, use and management easier and more human-centric, with the long-term goal of establishing full interoperability between operators.

We meet regularly and create publications to support operators, other MyData members, and the global personal data community. We seek to inspire the development of human-centric operator technologies, business models, and public policy that embody the MyData principles and identify collaboration opportunities.

Table of contents

Executive summary

This is an introductory paper to **MyData operators**: actors that provide infrastructure for human-centric personal data management and governance. An increasing number of businesses, legal experts, technologists, policymakers, and civil society actors are turning towards the general idea of approaching personal data use and management from a human-centric perspective. As our lives are increasingly digital, many of the rights we have secured in the physical world need to be carried over to the digital world. We also need to recognise new emerging rights and responsibilities native to the digital realm. In addition to laws and regulations, infrastructure for the management of personal data is also key to moving towards human-centricity in practice. The actors operating the infrastructure can guard the limits on what kind of activity is, and is not, possible or allowed or incentivised or rewarded.

The concept of **MyData operators** was introduced in the MyData white paper (Poikola, Kuikkaniemi and Honko, 2015) and the **MyData declaration** (MyData Global Network, 2017), but it has been empirically explored only at limited scale (European Commission, 2016; Lehtiniemi, 2017). We have taken the **MyData principles** as a starting point for this paper. We have studied existing examples of initiatives and organisations that are in one way or another either performing the role of an operator in personal data ecosystems or who offer related tools, services, or technologies. These **operators** can be considered **'trusted intermediaries'**. There is extensive literature and practice around trusted intermediaries of many forms and with many names: **infomediaries** (Hagel and Singer, 1999), **vendor relationship management tools** (Project VRM, 2008), **life management platforms** (Kuppinger, 2012), **personal data stores** (World Economic Forum, 2013), **personal information management services PIMS** (Ctrl-Shift, 2014), **personal information management systems** (Abiteboul et al., 2015), **information fiduciaries** (Balkin, 2016), **mediators of individual data MID** (Lanier and Weyl, 2018), **information banks** (MIC Japan, 2018), **data trusts** (ODI, 2018), **personal data co-operatives** (Hafen, 2019), or **data intermediaries** (European Commission, 2021).

The first edition of this paper (2020) was developed in collaboration with many potential **operators** existing at the time. It presented the 'state of common understanding' of what being a MyData operator entails. The paper was the reference for creating the **MyData operator** awards, and the key insights from that process are now captured in this second edition (2022). We present the **initial minimum requirements** to be considered a MyData operator in the paper. Common understanding and a shared language are essential for progressing towards the envisioned human-centric personal data infrastructure and ensuring interoperability between operators.

One of the central ideas of the **MyData operator** model is that there will be a large number of actors providing personal data management services. Those services should be interoperable and substitutable. Competing service providers should work together to create a global network for human-centric personal data transfer, similar to how different banks form a network for payments or mobile operators for phone calls. We recognise that this kind of interoperability is a journey where every step positively impacts people and service providers. The first edition of this paper, supported by the **operators**, provided the first step on this journey. Our ambition for the second edition is for it to build on the success and momentum to attract many more organisations to shape the future work needed.

Paper outline and research questions

The introduction describes the background to the concept of MyData operators as infrastructure providers in personal data ecosystems. We define ecosystem roles, what is expected from operators to demonstrate their adherence to the MyData principles and the idea of mutually interoperable operators. This is done based on the MyData declaration and other prior work.

We have gathered and analysed examples of over 40 potential operators from a dozen countries and engaged many of them in the process of compiling this paper. Our key questions when studying the landscape have been: *What are the functions a MyData operator should fulfil, and what responsibilities should it have? What is needed to create interoperability between the operators? What are the roles of legislation and governance frameworks in ecosystems, and how can operators bring better governance to human-centric data sharing? What are possible operator business models?*

These questions are addressed in the results section, where we present functional elements of the operators studied as a reference model and start defining multi-operator interoperability, human-centric governance and operator business models.

Reference model: The MyData operator reference model provides a structure to analyse operators' offerings and characterise their functional elements. The reference model creates a baseline for expectations for an operator from individuals, other operators, and other actors in the ecosystem.

Interoperability: Interoperability is key to realising the many benefits of the MyData vision. We describe different aspects of interoperability, recognising how these are currently prioritised by the operators and indicating the role of MyData in enhancing human-centric interoperability as ecosystems mature.

Governance: The governance of human-centric data sharing ecosystems is discussed in the contexts of legal and voluntary frameworks. We consider how governance should be formulated and enacted, taking into account transparency, the responsibilities of operators towards individuals and other stakeholders, and how the nature of who controls an operator impacts this relationship.

Business models: We study parameters of the business models options available to and currently used by trusted intermediaries, covering fundamental design criteria from the perspectives of human-centricity and financial sustainability.

The future work section addresses important questions raised during the work conducted for this paper, which deserves to be studied further. We conclude by summarising the MyData operator minimum requirements, and laying out a roadmap for progressing on the journey of interoperability with growing numbers of collaborating operators.

1. Introduction — MyData operators

Since the early days of the World Wide Web, the Internet has evolved from being a unidirectional broadcasting system, where companies showcased their products and services, to a multi-modal system with increased user engagement. This evolution has given rise to a situation where many technology giants have begun to track every action of every user with little or no transparency provided to individuals about the use of personal data so gained about them. Further, new business models have emerged based on selling data about people to third parties without consent from the individuals in question and with no means to opt-out. The result has been that clicking 'Agree' for consent was dubbed the internet's biggest lie (Obar and Oeldorf-Hirsch, 2018), and incidents of data misuse such as unsolicited calls, spam, and deliberate manipulation have resulted in a massive trust deficit.

Opportunities for innovation and efficiency have also been lost. The same data about the same individuals is collected repeatedly by every organisation that needs it, and this data is siloed and poorly maintained. Individuals cannot keep track of where data about them is held, and it rarely flows between platforms. And individuals are not the only ones to be harmed. The big technology platforms and large corporation systems now dominate markets to such an extent that many smaller companies, media organisations, and other market participants find it difficult to opt-out of these incumbent systems. Public actors, such as cities (Karhu et al., 2020), also face problems managing the personal data they collect, share, and use across their services or with contracted private actors. Public actors are not looking for ways to monetise data but need tools to process personal data ethically with their citizens in control.

1.1. Human-centric personal data

Organisations and initiatives are independently converging towards similar ideas about personal data infrastructure, management, and governance, where the people themselves would be in the driver's seat regarding the use and sharing of data from them and about them. This human-centric perspective promises to be the best and most inclusive approach to address the ills of the current data economy and, at the same time, to seize the opportunities for better use of personal data. By enabling individuals to share verified information with whom they wish when needed, we cut out the time, hassle and stress people experience when applying for services. At the same time, service providers' ability to rely on this information means they can streamline their processes. Both sides reduce their own friction, effort, risk and cost. Some examples of early communities focused on this topic include the Internet Identity Workshop[1], the Personal Data Ecosystem Consortium[2], and the Open Data & MyData Working Group under Open Knowledge Foundation[3].

Beginning with and facilitated by a series of international meetings and conferences from 2015 onwards, the MyData community has emerged as uniting supporters of the human-centric paradigm. The MyData declaration was published in 2017 as the joint understanding of the direction for MyData. The following year an international nonprofit organisation MyData Global was established. The human-centric MyData paradigm is aimed at a fair, sustainable, and prosperous digital society where the collective benefits of personal data are maximised by fairly sharing

1 https://internetidentityworkshop.com
2 https://pde.cc
3 https://personal-data.okfn.org/index-13.html

them between organisations, individuals and society. On the one hand, it seeks to ensure that people get value from data about themselves and can set the agenda for its use. On the other hand, MyData aims to establish the ethical use of personal data as always the most attractive option for organisations.

1.2. Ecosystems and infrastructure operators

Personal data is created, copied, moved, and used in ecosystems of individuals, data sources, data using services and actors in other roles. These ecosystems rely on infrastructure and infrastructure providers, who are crucially important in turning human-centric thinking into reality. There will always be at least one actor operating the infrastructure within any transaction that guards the limits of data processing by the actors involved. The MyData declaration asserts that this role must be carried out so that individuals can securely access, manage, and use the personal data about them, as well as control the flow of this personal data (MyData Global Network, 2017).

An infrastructure operator is positioned to connect the person and all other roles in the ecosystem. Besides operators, the viable use cases also need the participation of individuals, data sources, and data using services. If any one of these is missing, the case cannot exist. In business terms, the operators are in a multi-sided market position. The operators' value propositions should be viewed simultaneously from the perspective of individuals and organisations:

> For individuals: Operators provide transparency, understandability, and convenience to individuals when they share data or receive services using data about them. Operators support individuals to go beyond the control of their data to create their own uses for personal data and to re-use personal 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. When the use case doesn't include an external data source, the operator can facilitate transparency and act as a compliance mechanism for handling (collection or re-use) of personal data at the data using service. 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.

1.3. MyData principles for operators

While the MyData principles are highly aligned with data protection regulations in many countries and regions, they seek to empower people and communities with data, far beyond mere compliance with legislative requirements in any one jurisdiction.

The MyData declaration describes six principles for moving towards a human-centric vision of personal data. These principles imply the following requirements for relationships between operators, individuals, and other actors.

Human-centric control of personal data: This principle requires that any personal data transaction by an operator always involves[4] the individual. 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.

1.4. Mutually interoperable operators

Operators are not the end goal in themselves. Rather, they serve a role in the creation of sustainable and human-centric data infrastructures for personal data ecosystems. Different ways to organise personal data infrastructures exist and some of them are more aligned with the MyData principles than others.

It is easy to imagine at least four different high-level models for organising personal data infrastructures. These are not to be considered mutually exclusive, as there are multiple and differently organised, coexisting and overlapping personal data ecosystems, not a single overarching ecosystem. Also, within one ecosystem some hybrid of the below-mentioned models is possible (for example, a technologically decentralised ecosystem may have governance that is not fully decentralised).

4 Individual involvement may be setting preferences before the actual data transaction.

Fragmented: Markets where many small operator-like entities compete to build small-scale use cases without interoperability between them.

Monopolistic data platforms: A few platforms provide connectivity and data sharing inside their ecosystems with little competition and no incentives for interoperability between the platforms.

Fully decentralised: A peer-to-peer world where standardised technical infrastructure and protocols enable data connections without specific operator entities. In the decentralised model, the individual manages data flows directly from the end services or by having personal cloud-based applications on their own devices or hosted for them.

Competition-based interoperable operator network: Similar to the current network of telecom operators, energy providers, or banks where many mutually competing providers are interoperable and together provide global connectivity.

There is a common understanding that the first two scenarios (fragmented and monopolistic) are not desired states from the MyData perspective. It is hard to see human-centric principles sustainably maintained in them, however, they do describe the current starting point of the journey towards the more desirable scenarios (decentralised and competition-based). Many operators in the market are not interoperable yet but aim to be. Some operators may also evolve to build critical infrastructure for the decentralised scenario, for example, by integrating or embedding operator functionalities such as permission management seamlessly into service providers without the need for operators.

The ongoing debate over the relative advantages and disadvantages of the fully decentralised scenario is full of examples of both. Going fully decentralised may give developers the greatest flexibility to design or augment open-source software solutions that do not depend on trusting a third party. Technology should soon also allow self-sovereign peer-to-peer cloud storage. Counter-arguments for full decentralisation maintain that, even if technical infrastructure could be peer-to-peer, there are other reasons that operators would be beneficial as trusted intermediaries. The fully decentralised scenario could overly burden individuals with responsibility. Further, collective safeguards and regulatory oversight might be easier to establish in a model with clearly identifiable and possibly certified or licensed operator entities.

The competition-based interoperable operator network scenario would be comparable to telecom operators that provide global connectivity through shared standards and roaming arrangements. A mobile telephony system is far more beneficial for users than a fragmented system where one could only call phone numbers within the same mobile operator network. In such multi-operator networks, operators provide value to each other in addition to their value propositions to individuals and organisations. In an ecosystem with multiple mutually interoperable operators, this value is created from network effects and diminishing costs through collaboration, risk sharing and standardisation. Suppose each operator makes their connections to individuals, data sources, and services accessible to a common ecosystem. In that case, these operators collectively demonstrate a credible market and wide connectivity more quickly.

In the MyData community, there is general support for the competition-based scenario. However, the last two scenarios (fully decentralised and competition-based) can co-exist without compromising the MyData principles. This is possible if proper protocols exist for discovery and communication between the parties in the decentralised model and the operator network. In some cases, these two scenarios may even mix within the same offer.

1.5. From ecosystem roles to actors and functionalities

An ecosystem is composed of actors holding one or more of the main roles as described in the MyData declaration:

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.

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.

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.

Data Using Service: The role responsible for processing personal data from one or more data sources to deliver a service.

In practice, people and organisations do not get services from abstract roles; they get services from real-life actors. Different actors like governmental organisations, companies, and individuals can take the roles of an operator, data source, data using service.

In addition to the four roles above, originally described in the MyData declaration, we also recognise a role for **Ecosystem Governance**. This role is for actors that are responsible for managing, developing, and enforcing the governance frameworks for the ecosystem.

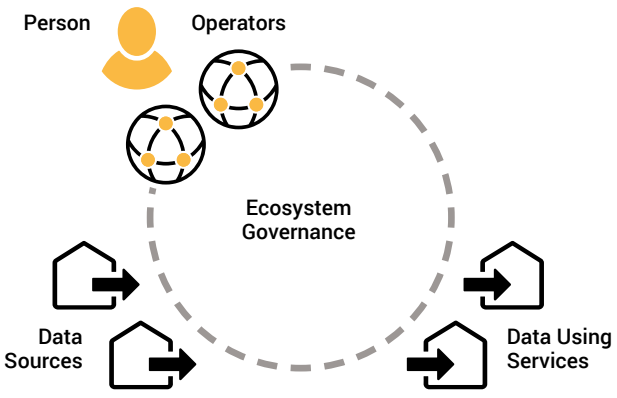

Figure 1: Illustration of a multi-operator ecosystem with the five roles of Person, Operator, Data Source, Data Using Service and Ecosystem Governance.

Example of a data transaction in a multi-operator ecosystem: A person in debt seeks debt counselling from their municipality. This debt counselling process can be supported by a specific operator for this purpose, which facilitates data gathering from multiple data sources (such as creditors, employers, tax authorities etc.) and the secure and controlled data transfer to data using services (such as the municipality and social security administration). It might even be the case that more than one operator is involved. For example, a specific operator focusing on health care costs may be used.

The role of an operator can have a wide range of functions associated with it. In this paper, we explore these functions and how those can be delivered in line with the MyData principles to understand the notion of a MyData operator further.

2. Methodology — studying the operators

This paper is the result of several years of work by many members of the MyData community. In the call for proposals preceding the 2019 MyData conference in Helsinki, several groups requested a workshop to explore the roles and definitions of the MyData operator. An open working group convened bi-weekly to prepare a briefing paper for the conference workshop (Janssen et al., 2019). Following the conference, the bi-weekly open calls continued to create the first edition of this white paper on understanding MyData operators by consolidating the learnings from the contributors.

In February 2020, the *MyData Operators Thematic Group* was approved by the MyData Global board to provide a structure for the ongoing initiative (MyData Global, 2020). The MyData Operators Thematic Group gathers a diverse range of individuals and organisations with long-standing experience in the interoperability and sharing of personal data. Many participants of the group run organisations that have operator functionalities are involved in the technical or service design of operator offerings. They have deep knowledge of how these functionalities are delivered across many sectors.

Working together, we compiled the list of 48 potential operators (proto-operators) from 15 countries shown in the appendices. The list was not exhaustive; rather, it was illustrative and reflective of the methodology of this paper. During our work on this paper, we approached organisations that we knew could qualify as operators. We requested them to read and comment on the paper draft and subsequently indicate if they wished to be included in the first edition of this paper.

Analysing the examples of operators collected, we saw a wide variety of actors in various stages of maturity with different technical approaches, business models, primary functionalities, offerings, and domains of activity. This diversity is a logical consequence of the early stage of the evolution of personal data ecosystems. It shows that the field is in a phase of rapid innovation and convergence, where standardised approaches are likely to emerge as the field's maturity grows.

In 2020, the first edition of this paper was published and a questionnaire was designed to allow operators to describe:

- how their approaches embody the MyData principles,
- the modularity of their systems with respect to the reference model,
- operation of their technical modules and associated interfaces,
- their business models, including data flows and value flows between actors.

This process of public self-description underpinned the MyData operator awards process on the basis of the sufficiency of the answers provided. The awards questionnaire was first run in 2020. It was then updated with improvements for the 2021 and 2022 MyData operator awards.

In the conclusion of the first edition of this paper, the editors and contributors suggested criteria by which potential operators would be awarded MyData operator status based on their self-descriptions (now shown below). These criteria were accepted by operators and informed the MyData operator awards process. The background for each of these minimum criteria is explained in the results section.

The initial minimum requirements to be considered a MyData operator are to **describe the systems for personal data management with respect to the MyData operators reference model** (interoperability), **demonstrate alignment with the MyData principles** (governance) and show that the operator business model satisfies the criteria of **transparency** with **the person as a primary beneficiary**.

This set of minimum requirements will evolve as the field matures. It has been possible to develop methods for the operators to self-describe as MyData operators based on these initial minimum requirements.

The operator self-descriptions from three years of awards and associated workshops within the MyData Operators Thematic Group have provided rich content that was previously unavailable. This material has allowed the editors to provide a greatly enhanced reference model description that forms the core of this second edition.

Our method throughout has been to uncover aspects of what is commonly understood among the various operators, constantly validating our findings with the group of contributors. What can be said about the state of this common understanding is more general than the state of the art of individual operators.

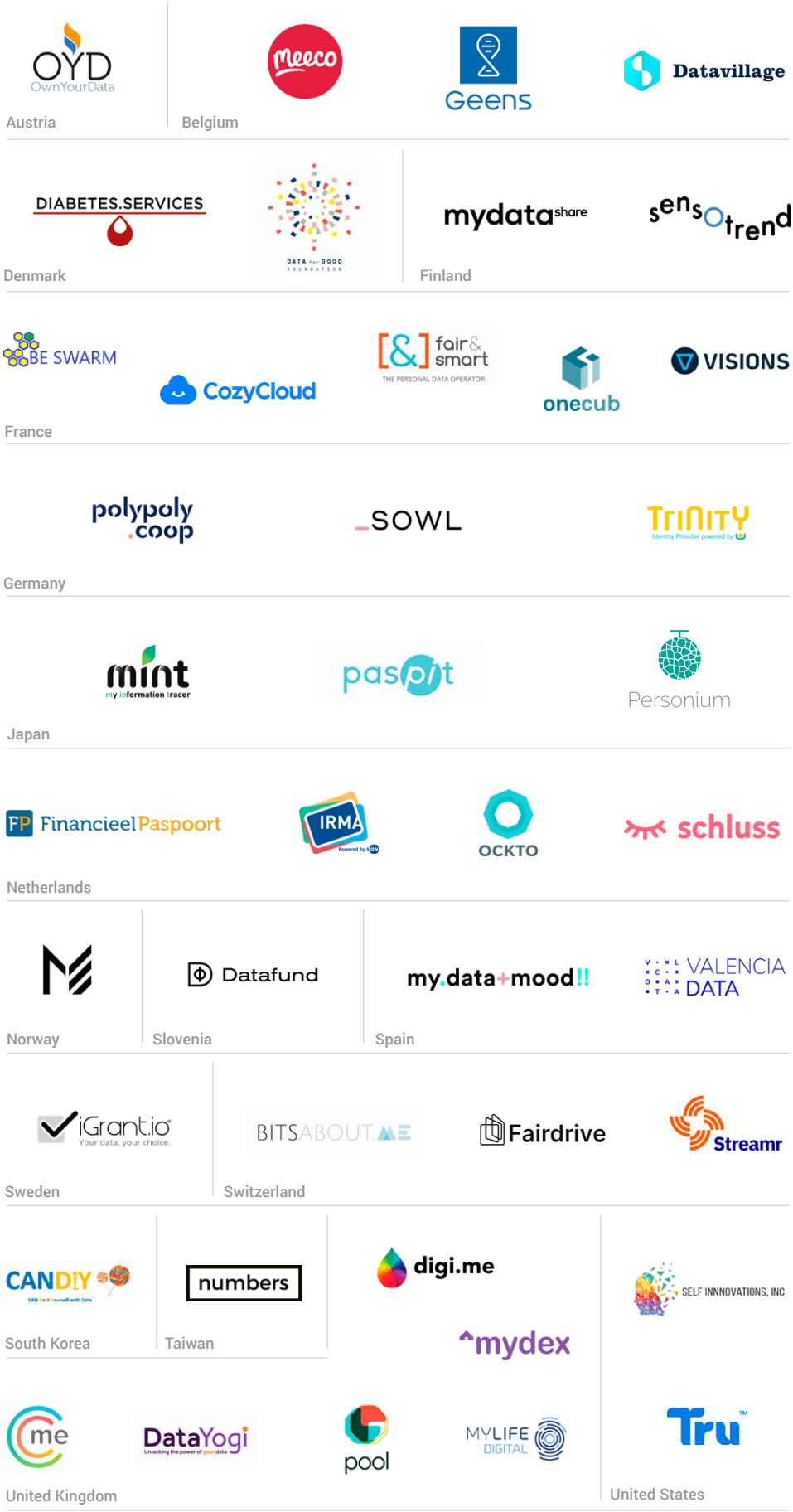

Figure 2: MyData operators awarded 2020-2022 (see Appendix 1: awarded MyData operators). This landscape will be updated regularly on the MyData operators webpage https://mydata.org/operators

3. Results — the state of common understanding

The results are presented under the broad categories of a reference model, interoperability, governance, and business models. These results have been derived empirically from our observations and analysis of the identified operators. Further, the results have been cross-checked and validated with these operators. They also reflect the current state of discussion in the MyData community. They are not intended to be normative guidelines but rather frame the debate to formulate more exact guidelines. At some future point, these guidelines may then contribute to binding rules.

This paper is not a call to tear down the offerings of the current operators but a challenge to make their functional elements visible to allow for digital rights to be exercised and to enable a fully functioning market. It is a call to action for those architecting new systems and applications using personal data to think clearly about who is performing the operator role and how they are empowering people.

The *MyData operator reference model* describes typical functional characteristics in many operators to understand better the commonalities and differences between operators. A crystal-clear picture of MyData operator archetypes is not immediately evident by studying the operators as they have different configurations of similar functionalities. The reference model has been structured to surface the differences of the operators studied, develop a shared vocabulary to discuss them and provide context for future harmonisation.

We use the metaphor of a 'journey of interoperability' throughout and lay out its initial roadmap with the *minimum interoperability criteria* for operators. At every stage in this journey of interoperability, MyData operators will be expected to assist people, in whatever way they can, to exercise their rights and to be empowered with their data. They must also always strive to work towards and within open networks, while innovating and creating differentiated offerings in a competitive market.

Balanced and fair relationships between people and organisations do not emerge automatically in personal data ecosystems. There need to be explicit human-centric governance methods to guarantee that MyData principles are followed. In the governance of *human-centric data sharing section*, we start to address questions regarding the extent to which an operator's responsibility is to 'create the balance' and guarantee human-centricity. And if it is not the operator's responsibility, then what other options are available?

Finally, the current state of *operator business models* is discussed with outlined design criteria for future operators. As the underlying business model strongly influences operators' functions and modes of activity, it is important to define what kinds of business models are aligned with the MyData principles and which models might not be.

Under the interoperability, governance and business models sections, we present the initial minimum requirements to be considered a MyData operator.

3.1. The MyData operator reference model

The MyData operator reference model describes nine core functional elements of operators. These elements affect how easy it is to utilise personal data, how transparent and human-centric the utilisation of personal data is, and how well the infrastructure supports open competition.

In the rich and complex landscape of operators, a basic common understanding of the types of functionalities offered is needed to transition from a fragmented landscape of solutions to sustainable personal data ecosystems. This reference model is a tool for the operators to describe their functionalities using shared terminology and collectively advance the interoperability by gradually converging to common standards.

This reference model results from iterative synthesis from studying the wide range of functions that existing operators currently support. All the elements described are present in many of the operators, and they are commonly considered important or even essential for realising sound and sustainable personal data ecosystems. The empirical understanding gained from our research has been validated against previously presented conceptual models of key technical solutions for human-centric personal data management (Poikola et al., 2015; Rikken et al., 2019; Sitra, 2020).

We acknowledge that important properties, such as information security, are not included in this model as they are general requirements to all personal data services. We have selected elements for inclusion in the reference model based on the criteria that they are relevant in the context of MyData, help to differentiate between operators, and are directly valuable to individuals. Also, using data as the means of payment and directly paying individuals for their data are contentious issues. The MyData operator reference model does not intend to solve or move that debate in any direction.

The reference model should not be thought of as a monolithic template for direct implementations. We emphasise that not all reference model elements need to be part of all operators. Value exchange, for example, may not be an important aspect in many ecosystems but can be essential in commercial settings where it needs to be implemented according to the MyData principles. Functionalities can also be distributed or duplicated over the different roles in the system: not everything resides with an operator in isolation, and some functions might apply to all roles (e.g., logging). How operators choose which elements to support falls outside the scope of this paper. To help understand the elements in the context of each other, we include a connection matrix below that indicates how the different elements work together and are connected.

Technologies and standards related to each one of the functional elements are being developed independently of each other and independently of MyData. When describing the functional elements, we deliberately do not reference any particular technology or standard as we acknowledge that today's technology choices are subject to change in the future. Evolving legislation, technology, standards, and organisations operating in personal data management all affect how operators eventually implement the functionalities presented in this reference model. We include the first landscape of standards and specifications most commonly in use by the current MyData Operators for general guidance.

The elements of the reference model

The set of functional elements that a MyData Operator implements influences the functionalities that an operator can offer to the individuals and other entities in the ecosystem. Each of the nine core functional elements enables some unique capability that is not covered by any other element. However, the elements should not be considered in isolation as most of the functionalities that are valuable to the other parties in the ecosystem involve more than one functional element.

A summary of the unique capabilities of the functional elements follows below after the image. We then present a matrix table showing the relationships between the functional elements and how they impact the human-centricity of the ecosystem. A more detailed description of each element follows.

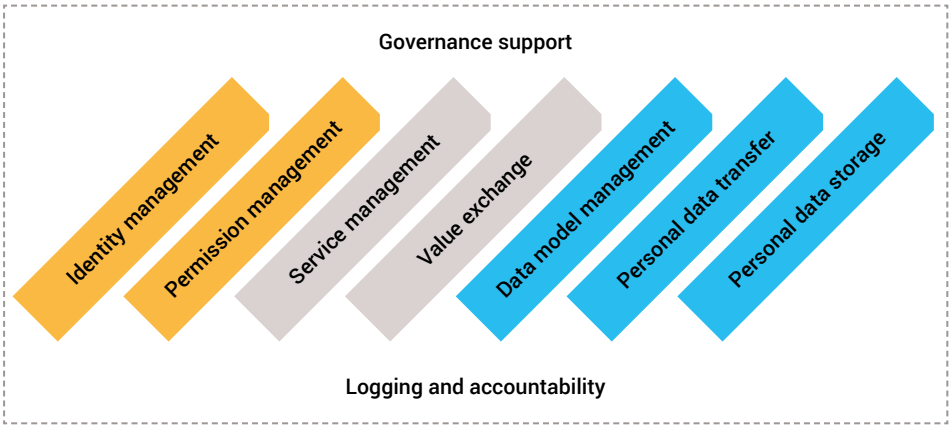

Figure 3: Functional elements of a MyData operator. The first two (yellow) pillars mediate data transactions in terms of participants and permissions. The middle two (grey) pillars describe what services are enabled in the ecosystem and how value can be exchanged between ecosystem participants. The right-hand three (blue) pillars manage data, its meaning, its exchange, and its storage. 'Governance support' and 'Logging and accountability' provide context for the other functional elements and are critical for transparency and trust in the ecosystem.

Unique capabilities provided by the functional elements:

Identity management (IM) functional element enables the identification and authentication of the different actors involved in data exchange.

Permission management (PM) functional element enables the configuration and authorisation of data exchanges that are allowed to happen (e.g. notices, consents, permissions, mandates, legal bases, purposes, and preferences), and collection and handling of such authorisations.

Service management (SM) functional element enables maintenance of a register of known actors and information about them (e.g. supported data models, preferences, and discovery information) in the data ecosystem.

Value exchange (VE) functional element enables accounting and the transfer of value (monetary or other forms of credits or reputation) related to the exchange of data.

Data model management (DMM) functional element enables data harmonisation and conversion into specified formats, communicating semantics (meaning) of data with other ecosystem participants and interpreting standard data models to individuals.

Personal data transfer (PDT) functional element enables data exchange between the ecosystem participants in a standardised and secure manner by implementing permissioned interfaces (e.g. APIs).

Personal data storage (PDS) functional element enables data integration from multiple sources (including data created by a person) in personal data storage under the individuals' control and serving data from PDS to data using services.

Governance support (GS) functional element enables compliance with the underlying governance frameworks to establish trustworthy relationships between individuals and organisations.

Logging and accountability (LA) functional element enables the maintenance of records (including record deletion) on data exchanges taking place and creating transparency about who accessed what and when and based on what permissions.

Matrix of functional element relationships

Table 1: The cells in the interaction matrix describe each relationship between the functional elements (rows and columns).

	Permission management	Service management	Value exchange	Data model management	Personal data transfer	Personal data storage	Governance support	Logging and accountability	
Identity management	The identities of the parties are needed to codify who is granting permissions to whom.	The identity of a service is needed to allow discovery in service management.	Identity management broadly defines who is eligible or accountable in value exchange.	Identity management can use transaction data models for describing identity in detail.	The identities of authorised participants are needed for personal data transfer.	Authenticated identities are needed to verify access and control for personal data storage.	Governance policies define what kind of identity management is required for which action.	Identity management can implement logging of identifiers and authentication events.	**Identity management**
Permission management		Service management provides contextual information about the services to which permissions may be granted.	Value exchange enables the transfer of value related to permissioned data sharing.	Transaction data models define the structure and granularity of permissions.	Technical enforcement of permissions can control the transfer of personal data.	Access control and authentication mechanisms are needed for permissioned data storage and access.	Governance policies can be encoded into permissions.	Logging modifications of permissions ensure that the parties can be accountable for following the permissions.	**Permission management**
Service management			Value exchange rules are controlled within service management structures.	Service management can use machine-readable data models for service descriptions.	Provides discovery of interfaces for data sources, models and functions used in personal data transfer.	Personal data storage should be addressed as a service or group of services within service management.	Governance policies can be applied through service management when vetting and onboarding services.	Service management can implement logging of service registration and modifications.	**Service management**
Value exchange				Transaction data models help to define metrics for value exchange.	Value exchange is based on permissioned data transfers.	A PDS containing wallet functionalities can store value (tokens).	Governance policies can define if, when and how value exchange can be realised.	Logging can be used either as a measure for value exchange or audit for valid billing for services or data.	**Value exchange**
Data model management					Semantic and transaction data models harmonise personal data transfers.	Semantic and transaction data models (including standard APIs) enhance the utility of personal data storage.	Governance policies can be specific to data model or require linkages to data semantics.	Transaction data models support interoperability of logging and accountability.	**Data model management**
Personal data transfer						A personal data store can be a source and/or a target in personal data transfer.	Transaction related governance policies can be applied through personal data transfer functionality.	Logging of personal data transfer activities forms part of a verified transfer documentation.	**Personal data transfer**
Personal data storage							Governance support defines the role of personal data storage in the data exchange environment.	A personal data store can log all transactions to provide transparency to the person.	**Personal data storage**
Governance support								Logging provides records for auditability of compliance with the governance policies.	**Governance support**

Human-centricity in the reference model

The system's usability and individual's control define the human-centricity of an **identity management** system. The concept of self-sovereign identity SSI (Wang and De Filippi, 2020) can have an advantage over federated and centralised identity models as the locus of control is more directly with the individual. The understandability of any identity model needs to be carefully addressed to be more human-centric.

Value exchange can only be human-centric when the person is involved in or aware of all stages and relationships in the exchange. The person must be a primary beneficiary of an exchange involving an operator.

The human-centricity of **personal data transfer, permission management** and **service management** depends on the granularity of control and the understandability of the interfaces that provide the control. These need to be usable and intuitive while also being fine-grained and configurable. They need to provide overviews of both transactions and relationships. The interfaces and controls must not be a bottleneck or the weakest link.

A key contributor to the human-centricity of **personal data storage** is the usability of key management and levels of encryption. Technically, personal data storage should be implemented so that an operator does not have access to the data and has little to no information about what data the person stores. **Data model management** should help the person interpret the data about them, including personal and transaction data.

Logging and accountability should provide the person the ability to control logging within the limits of the governance framework and regulatory requirements. Visibility, transparency and data minimisation (restriction of excessive logging) must be balanced for all parties to achieve optimal accountability.

Identity management

Identity management (IM) *functional element enables the identification and authentication of the different actors involved in data exchange.*

Managing the identities of individuals and confirming the identities of other actors in the ecosystem makes it possible for individuals to act as the 'point of integration' regarding data about them. Individuals can have different identities, or profiles, with different data sources and data using services. For example, they can have public and private identities or self-sovereign identities. The role of identity management within a MyData ecosystem is to assist actors in making informed decisions about who they want to interact with. This applies to individuals, organisations and services – all need to be identified and authenticated to enable trusted data exchanges. This trust is built, in part, on providing the appropriate level of assurance in each instance.

A functional operator service without some identity management is impossible as authenticated identities of parties are needed to achieve any effective data flows. A MyData operator service can use external identity providers and implement only lightweight identity management to do key matching between the data source and the operator (also known as 'blind linking'). On the other hand, an operator can also offer a full identity management service.

MyData promotes human-centricity through individuals being able to manage their accounts and personal data, giving them the ability to interact in the digital world with the same freedom, pseudonymity, and capacity for trust as they interact with organisations in the offline world. The concept of self-sovereign identity SSI is well-aligned with human-centric personal data management and a promis-

ing approach to overcome verification and privacy issues and enable interoperability across the ecosystems in a scalable manner. In the self-sovereign identity model, the individual can control of their digital identities. SSI addresses the difficulty of establishing trust in an interaction. To be trusted, one party in an interaction will present credentials to the other parties. Those relying parties can verify that the credentials came from an issuer they trust. In this way, the verifier's trust in the issuer is transferred to the credential holder. This basic structure of SSI with three participants is sometimes called "the trust triangle".

Some operators also copy ('cache') identity attributes, allowing them to function as a log-in tool. There is a grey area between operators and authentication-only tools, as some data attributes are also exchanged when identifying a person during authentication processes.

Uniquely in the scope of this element:
Identity management uniquely covers the authentication of actors in the ecosystem. MyData operator can realise this using different identity management models and related authentication protocols:

- **Centralised identity management:** based on a proprietary or external identity provider (IdP) supporting authentication protocols such as LDAP or OAuth2. It may use an API key for enabling service- or application-level authentication, while individual to application authentication may use an HTTP authentication schema (e.g. a Bearer JWT token).
- **Federated identity management:** supported with shared protocols such as OpenID connect or SAML authentication. It may also provide out-of-the-box support towards any existing identity stack (e.g. national bank ID or eIDAS). Note that this is only applicable to individuals as there are currently no large-scale ID federation schemes for organisations.
- **Self-sovereign identity (SSI):** Self-sovereign identity (SSI) is an approach to digital identity that gives individuals control of their digital identities. In an SSI system, identity holders generate and control unique identifiers called decentralised identifiers (DIDs). Verifiable Credentials standard (W3C recommendation) is an open protocol for issuing, holding, and verifying digital credentials. Data Sources can issue verifiable credentials tied to a pairwise DID that the individual can use to make claims towards a Data Using Service without any interaction with the Data Source. A DID is associated with a DID method via the DID Document that can be resolved independently by the Data Using Service or any third party via a DID registry. The DID Document contains the public keys associated with the DID that can be used for any verification, e.g. authentication. DID authentication refers to a method of proving control over a DID for the purpose of authentication. Self-Issued OpenID Provider (SIOP) is a flavour of DID authentication to use OpenID Connect (OIDC) and strong decentralisation.

Positioned between federated and self-sovereign identity systems are user-centric identity approaches that provide a life-long digital identity that can be used anywhere but is not decentralised. These systems are often sector or state-defined and limited by those boundaries.

Key, questions assessing the identity management capability of an operator are:

- Is a "full" IM service provided or only lightweight key-matching using another IM service?
- Which identity management models (see above) and related authentication mechanisms does the operator support?
- How does an individual access the operator service? How do they log in?

Scales of Interoperability:
Interoperability for the identity management functionality depends on the support for shared authentication protocols among the operators.

Delivered with other functional elements:

- The identity of data sources and data using services remains uniquely in the scope of identity management functionality. Still, service **management** covers the discovery of these services and enriches the information about the services, for example, by validating the legal entities against the public business registries.
- **Personal data transfer** relies on identity management for data sources to prove the person's identity whose data will be accessed.

Permission management

Permission management (PM) *functional element enables the configuration and authorisation of data exchanges that are allowed to happen (e.g. notices, consents, permissions, mandates, legal bases, purposes, and preferences), and collection and handling of such authorisations.*

The term 'permission' is used in a broad sense to cover the mechanisms used by the individual to control the use of data, and the data flows. These mechanisms can be based on legislation (executing legal rights such as consent) and more detailed preferences.

Several operators focus primarily on permission management, providing a way for people to orchestrate which data is shared for whom, for which purposes, and for how long. These operators often have a core value proposition of facilitating legal compliance for the data sources and data using services involved in the data transactions.

Existing and emerging standards (see Appendix 3) provide interoperable data structures for handling permissions. This includes the data usage purpose, the legal basis of the use of data, and the retention period. These standards enable interoperability where individuals can share their permissions from one operator to another. The standards also allow a consistent user experience across different permission management services.

Implementations of permission management may enable different types of permissions:

- **Content permissions:** marketing, offers, customer research, newsletter, membership etc.
- **Data sharing permissions:** permission to share data onwards within a group or with partners etc.
- **Data use and purpose permissions:** permission to process data for a particular purpose such as automated decision making, tracking location or online activity, selling personal data etc.
- **Data portability:** e.g. permission to move data to a different country.

The types of permissions mentioned above have reference points in regulations such as the GDPR, and the list will evolve. Building a standardised list of permissions and agreeing on the naming and design of data attributes will advance the interoperability of the operators. The permission types listed above often overlap in the real world use cases where several types of permissions may be asked and granted in a single transaction.

The different types of permissions can be encoded in data sharing agreements that all parties of the data transaction agree and comply with. The data sharing agreements should follow all relevantvernance policies (see governance support) and be generated using appropriate standards in human and machine-readable formats.

Uniquely in the scope of this element:
Permission management covers the technical functionalities required for human-centric control of personal data, such as the user interfaces and underlying data structures for individuals to view, understand, grant, revoke, and modify different permissions related to data flows. It may provide the ability to control permissions at the level of data usage purpose and with further granularity at the attribute level. This functional element also technically enables enforcing permissions so that an operator only allows personal data transactions based on valid permissions.

The permission management functional element enables the generation of data sharing agreements based on contract templates or clauses from a policy register, where relevant. Further, the functional element enables providing the agreements and related metadata (parties, data set, policies, time) to all parties of the agreement. The authorised data ecosystem participants can check the existence of a permission, get relevant metadata about the permission and a token proving the validity of the permission that can be used to enable data exchange. Generating new data sharing agreements and providing information about existing agreements could be manifested to the authorised ecosystem participants via a contract API.

Key questions assessing the permission management capability of an operator are:

- What types of permissions (see the list above) does the operator support?
- How does the operator ensure valid permission (consent or other) for data exchange between a data source and data using service? What protocols are used?

Scales of Interoperability:
The permission definitions (receipts) and the identifiers needed to identify parties of given permission should be universal enough to allow an individual or organisation to switch freely from one MyData operator to another (permission portability). At the level of user interfaces, the implementation of permission management should be universal enough to allow individuals to share data correctly using different MyData operators.

Delivered with other functional elements:
Permission management is central to operator services and connected to several other functional elements:

- **Identity management** and **service management** can be used to codify permissions (i.e. determine the different parties of the data transactions and attach additional information such as the end-points from the service registry).
- Permission management technically enforces the permissions so that **personal data transfer** is allowed. Permission management may also support use cases where a data using service requests additional data processing permissions, via an operator, for data to which it already has access for existing purposes.
- **Governance support** sets requirements for permission management derived from the legislation and ecosystem-level governance frameworks.

Service management

Service management (SM) functional element enables maintenance of a register of known actors and information about them (e.g. supported data models, preferences, and discovery information) in the data ecosystem.

Operators live in an ecosystem with data sources and data using services. Navigating this ecosystem requires linking actors through an operator: this is the purpose of the service management functionality. The human-centric manifestation of service management is the possibility for individuals to manage the relationships and connections to different data sources and data using services in the ecosystem.

Service management enables dynamic linking of data sources and data using services (permissioned by the individual) so that data can be available at different sources and can be used by multiple data using services.

It is a significant decision in a multi-operator environment if the operators use a shared service registry (potentially still distributed) or if each operator manages services separately. This topic will evolve in future work; currently, there is limited standardisation or convergence in this field.

Uniquely in the scope of this element:
The central components of the service management are **service registry** and **relationship management.** The service registry enables the discovery of trusted data sources and data using services that individuals can then link to with the relationship management tools.

The service registry responds to the question: "what does an operator know about the data sources and data using services in the ecosystem?" The register could contain, for example, the basic information (address, representatives, etc.), types of data available in the sources, supported data models or access endpoints. Besides technical information, the register can also contain information on the legitimacy of the services, for example, by validating the legal entities against the public business registries and maintaining information on how these services have been vetted when onboarding to the data ecosystem.

The relationship management responds to the question of which data sources and data using services the individual is connected to and offers tools for the individual to manage these connections. The relationship information should be accessible with a standard request interface so that the data using services can query from an operator what data sources the individual has already as linked connections and

request an individual to add new connections. The visibility of the individual's connections to the data using services must be under the individual's control.

Key questions assessing the service management capability of an operator are:

- What kind of service registry does the operator use?
- How is information about the services maintained in the service registry?
- How can individuals manage connections to data sources and data using services via the operator?
- Does the operator offer architecture to facilitate harmonised access to data sources?

Scales of Interoperability:
Operator interoperability (enabling substitutability and enhancing roaming possibilities) in service management would be advanced by developing interoperable service registries or a shared distributed service registry. Another dimension of interoperability is standardising interfaces for data sources with architecture supported by the operators.

Delivered with other functional elements:

- Service management and **data model management** combine to offer standard interfaces for the data sources. Although data sources may largely determine the delivery of the interfaces, the operators can support the harmonisation of these interfaces by offering a standard connector or socket architecture and even help data sources to get onboarded and implement such connectors or sockets.
- Service management needs **identity management** because the data sources and data using services need identities.
- Vetting and onboarding (and expulsing) of data using services and data sources to the data ecosystem technically happen via the service management functionality, but it shall comply with the governance policies (audit, contracts, regulatory compliance) stemming from the **governance support** applied.
- Service management can also support the orchestration of end-to-end transaction flows across multiple actors, including individuals, enterprises or IDPs. Such orchestration of multi-party transactions typically involves **identity management** and **permission management** and may also involve other functional elements.

Value exchange

Value exchange (VE) *functional element enables accounting and the transfer of value (monetary or other forms of credits or reputation) related to the exchange of data.*

Value exchange enables business models – it allows the creation and distribution of value in the operator ecosystem. An ecosystem business model is a blueprint of how a network of cooperating organisations intends to create and capture value from services or products (Haaker et al., 2006). In the case of MyData operators, from facilitating the data exchange or exchanging or processing personal data. Some MyData operators may also support value exchange, which is not directly related to data transfers, e.g. payments or non-fungible tokens (NFTs). However, value exchange beyond data is not in the scope of the MyData operator reference model that focuses on the operator functionalities.

Sustainable business models are a requisite for ecosystems in general. Business models are about creating value, and all participants in the ecosystem need to have more benefits than costs in the long run. Value, benefits and costs can also be non-monetary (also, not-for-profit organisations have a business model). For individuals, time and effort can be a high cost, and benefits often come from the services.

In total, a personal data ecosystem that adheres to the MyData principles should create more value than the overall costs incurred by the participants collectively. However, value creation does not happen equally in all parts of an ecosystem and mechanisms for distributing value are needed.

To identify the value exchange in an ecosystem, we can follow the flows between the different actors in the ecosystems. These flows can be data, services granted or accessed, or even societal value flows. There should be reciprocity of flows between the actors (Gordijn, 2004). If there is a data flow from a data source to an operator, there should also be a value flow in return from operator to data source – for example, compliance to GDPR.

Uniquely in the scope of this element:
As operators provide technical infrastructure for making multi-party data transactions possible, they are also in a natural position to keep track of such transactions for payments and billing or creating other forms of rewards, such as loyalty and bonus points. Operators may provide a value 'accounting' mechanism that transparently keeps a log of the data transactions. The different parties in the ecosystem may use it as the base for payments. Also in the scope of this functional element are operator enabled payments or other types of explicit value transfers besides the transferred data.

Key questions assessing the value exchange capability of an operator are:

- What kind of value accounting mechanisms does the operator provide?
- What types of explicit value transfers does the operator enable?

Scales of Interoperability:
Interoperability for the value exchange functionality depends on the trustworthiness, wide recognition and fungibility of the value accounting mechanisms it supports.

Delivered with other functional elements:
Every one of the functional elements has the opportunity to deliver value through new benefits and improved efficiency. The configuration of functional elements for a specific operator needs to be optimised to maximise the overall value created at a minimum cost to give a sustainable business model. Promoting services will often emphasise the value delivered in a certain area to provide market differentiation.

Data model management

Data model management (DMM) *functional element enables data harmonisation and conversion into specified formats, communicating semantics (meaning) of data with other ecosystem participants and interpreting standard data models to individuals.*

Data models aim to define the terminology used across processes, services and data, including the semantics, structure and format of data structures. Harmonisation of personal data models strengthens options and potential for data portability and increases data usability. Without standardising the definitions of used concepts, each data source and using service will require and develop its own definitions. Not only is the lack of standard definitions costly, but multiple definitions can also lead to confusion over the relationship between data assets and the real-world concepts they describe.

Designing a data model starts from scoping the systems that the data model would be required to support. For example, the data model required for a carbon calculator app would differ from that needed for a smart home dashboard. The scoping discussion is followed by defining the data elements' structure and relationships and, finally, technical implementation specifics.

Generally, the data model management can be thought of in three phases:

1. The scoping discussion delivers a **conceptual data model**
2. The **logical data model** defines the structure of the data elements and the relationships between them
3. The **physical data model** specifies a technical implementation

MyData operators work in different domains, each having business-specific technical requirements related to data models. Developing, harmonising and maintaining data models is tied to real-world use cases. Standardising the physical data models (phase 3) across all different MyData operators is unrealistic. Depending on the domain, semantic data standards are more or less evolved. Until widely adopted standards exist, translations between data models are necessary.

As MyData operators support the management of personal data, one can assume 'person' will always be an object within the data model. Therefore, the used data models should have a significant commonality at the core. It is possible to build common views to inform human-centric data models (phases 1 & 2).

Some operators offer data harmonisation as a service, while others focus more on the data transfer leaving the data model management for the data sources and data using services. Personal data management without data model management is possible but limited in scalability, interoperability, and data usability. Also, if a data using service depends on a specific type of data harmonisation uniquely provided by one operator, it may become a barrier to change operator.

Transaction data models and semantic data models
It is important to differentiate the **semantic data models** related to the content itself from the **transaction data models** that enable the participants of the data ecosystems to share and manage data. Transaction data models define, for example, the structure of identity claims, permissions, service definitions, standard APIs, governance policies and log data syntactics. Harmonisation of the transaction-related data models is crucial for the interoperability of the data ecosystem. Therefore, operators that do not support semantic data model management should still support some transaction data models.

Uniquely in the scope of this element:
Data model management as an operator functionality covers tools and services to facilitate translation of one data model to another, master data management and data governance within the data ecosystem.

As many data standardisation processes are not human-centric, data model management as an operator functionality can also facilitate interpreting standard data models to individuals. Supporting commonly agreed human-centric data models is also in the scope of this functional element[5].

Key questions assessing an operator's data model management capability are:

- What data models does the operator use internally?
- What standardised data formats will the operator accept?
- What tools for data modelling, management, and governance does the operator offer?

Scales of Interoperability:
Interoperability for the data model management derives from an operator's support of the human-centric data model and commonality of the offered tools and services.

Delivered with other functional elements:
The transaction data models (non-content-related data models) are not uniquely in the scope of the data model management functional element but are delivered with the respective functional elements. For example:

- **Identity management** needs data models for authentication and verification.
- Data sharing contracts and permission receipts are data models related to **permission management**.

5 The MyData dictionary https://mydata.org/mydata-dictionary is a data model built from the perspective of the individual, not from that of organisations. It includes the key data fields an individual would expect to see in a data-set.

Personal data transfer

Personal data transfer (PDT) *functional element enables data exchange between the ecosystem participants in a standardised and secure manner by implementing permissioned interfaces (e.g. APIs) that rely on operator/s for controlling access to the data.*

Personal data transfer, through an operator or facilitated by an operator, is a key concept in leveraging value from personal data as it technically enables portability, access and re-use of data.

A MyData operator can technically facilitate personal data transfer with different models:

1. data can flow through an operator;
2. an operator can facilitate the direct transfer from data source to data using service under valid permission.

In the first case, an operator implements the interfaces for data transfer (uniquely in the scope of this element). In the second case, an operator caters for permission management but does not implement the data transfer interfaces.

In both cases, operators need to manage the transfer of personal data in line with permissions (see permission management) and ensure compliance with the governance framework reflecting applicable laws and the individual user's needs (see governance support). The applicable governance framework may set limitations to data transfers regarding the physical location and jurisdiction of the receiving end of the data transfer.

'Data sharing' is a catch-all term that is often used in the context of personal data transfer and hides a multitude of variations. The user may send data to a data using service to gain new insights but requires privacy-preserving mechanisms not to leak any information. The user might also share access to data to earn rewards, including improved services or monetary compensation. The user may decide to donate data or transfer it to be managed by a data trust. These are just a few examples of the reasons why data might need to be transferred. However, the data transfer functionality is agnostic to why data is transferred in the first place.

Uniquely in the scope of this element:
This functional element enables data movement in a standardised and secure manner between the ecosystem participants. These participants can be personal data sources (e.g. smartphone, personal data store) and third-party data sources, data using services (e.g. data processor, storage provider) and other operators (roaming in an operator network). While data transfer is about moving the data and not storing it (see personal data storage), it can still use transient data stores to facilitate the transfer. The data transfer functionality should control for the removal of the data after such transient handling.

Key questions assessing an operator's personal data transfer capability are:

- What dynamic (e.g. time series from sensors) and static (e.g. files, credentials, etc.) data transfer types do the operator support?
- What types of interfaces (e.g. REST API) and connections (push, pull) and possible secure and privacy-preserving processing capabilities the operator enables for data transfers?

Scales of Interoperability:
The interoperability for the personal data transfer functionality derives from supporting common standards on the aforementioned aspects (types of data transfers, interfaces and connections). The weight of each aspect needs to be considered based on the user needs and requirements in a particular ecosystem.

Delivered with other functional elements:
The MyData operator may facilitate data transfer with other functionalities such as discovering data sources and data using services, authentication of the transfer parties, logging, provenance of data sets and data model transformations. Operators may also support data governance and management to ensure that data is not unnecessarily duplicated and can be updated easily across any copies when required. These highly valuable compound functionalities require the involvement of other functional elements and are therefore not unique to personal data transfer.

- The questions of what data is shared, with whom, for what purposes and based on what permissions are generally in the scope of **permission management**, except for what is required by the technical requirements for the data transfer interfaces.
- Standard interfaces for the data sources ('socket' or 'connector') and discovery of data models and functions used in personal data transfer are in the scope of **service management** and **data model management**.

Personal data storage

Personal data storage (PDS) *functional element enables data integration from multiple sources (including data created by a person) in personal data storage under the individuals' control and serving data from PDS to data using services.*

Besides a functional element provided by an operator, a PDS can also be considered a separate data source and/or data using service controlled by the individual. It is included in the operator reference model as it is a central part of the offering of many operators. In practice, the operators are well-positioned to offer a PDS.

Using PDS as a 'station' for personal data configures the connections in the data ecosystem so that the different parties of data transfer can connect via the person and do not need to be directly connected to each other. This configuration may also simplify legal liabilities and the implementation of permission management.

Enforcing the separation between data sources and data using services is a potential path to increase human-centricity in the data ecosystem. The person with a physical or virtual PDS is technically in the centre of the data transactions. Operators offering PDS solutions strive that people would hold up-to-date 'personal master data' for commonly used attributes and data types, such as contact and preference profiles. This reduces the need of having the data duplicated (and often outdated) in many places.

Are digital wallets in the scope of the personal data storage functional element?
A digital wallet is an application that allows the individuals to store and manage identity data, credentials and attributes linked to their identity. The stored credentials can be used to sign transactions, statements, documents, or make claims. A digital identity wallet enables an individual to establish relationships and interact with third parties in a trusted manner.

Personal data storage (PDS) functional element of the MyData operator reference model does not cover the identity- and key management functionalities typically related to wallets as these functions are defined in the identity management element. However, some wallets extend from key- and identity management towards management of personal data by supporting storage of verifiable credentials and other types of personal data. Therefore, a wallet application may also implement the PDS function.

Uniquely in the scope of this element:
Personal data storage (PDS) functionality allows stored data under the individual's direct control to be integrated from multiple sources — harmonising, using and re-sharing it.

As PDS is also a data using service, it can be extended via plugins or apps running next to the personal data store. Such plugins may enable various operations for the data and with the data, including data cleansing, harmonisation, integration, analysis etc. The extensibility is uniquely in the scope of this functional element, even if the individual plugins are not.

Key questions assessing an operator's personal data storage capability are:

- Where and how the data is hosted (on device, on cloud, on server)?
- What is the level of decentralisation (centralised, decentralised, self-hosted) of storage?
- What storage formats the PDS supports as part of data model management (files, flat (e.g., JSON), graph) and if there are differences to what you get from the API as part of personal data transfer?
- What kind of data encryption is supported (not encrypted, symmetric, asymmetric), and how does the encryption key management work?
- Is the PDS extensible through plugins or apps?

Scales of Interoperability:
The scale of interoperability for the PDS functionality is defined by the storage standards, hosting options, and the universality of supported interfaces for apps, plugins and extensions.

Delivered with other functional elements:
Standard APIs, permission control and authentication mechanisms, as well as the audibility of the PDS, are integral to a trusted and functioning PDS. However, they are not uniquely in the scope here as they require the involvement of other functional elements.

Governance support

Governance support (GS) *functional element enables compliance with the underlying governance frameworks to establish trustworthy relationships between individuals and organisations.*

Human-centric governance helps to mediate the relationships between people and organisations. This dedicated functionality in an operator can guarantee that My-Data principles are followed and enable compliance with underlying governance frameworks.

To some degree, all operators operate within a governance framework to be transparent about the trustworthiness of their services. Operators may be able to select governance frameworks within which to work, or they may have to respond to mandatory requirements within their sector and jurisdiction.

All data transactions within a governed data ecosystem follow some rules and conditions, such as pre-set policies or dynamic conditions specified by the transaction participants. These rules and conditions regarding data use and sharing are collectively called **governance policies**. They may cover: *codes of conduct, obligations, restrictions, prices, terms of service, certifications, data protection and security requirements, rights, liabilities etc.*

Establishing governance bodies or setting and changing governance policies are not in the scope of the operator reference model. These questions are the primary concerns of the ecosystem governance frameworks discussed later. The governance support element contains the functional counterparts on the operator level to support such ecosystem governance frameworks.

Technical realisation of governance

Governance policies translates into responsibilities for an operator, which can then result in liabilities in a well-governed ecosystem. The governance policies state what different parties 'should do' and can be 'expected to do', of critical importance are the actual governance practices, 'what is done'. Auditing provides the mechanism to reconcile any differences between policy and practice.

Technically, an operator implements the governance policies with many functional elements. The role of the governance support is to maintain up-to-date governance policies from external sources (e.g. the governance framework) and serve the policies so that the different functional elements can implement and enforce them in practice. For example, the permissions generated with the permission management functionality can link to the machine-readable governance policies served by a policy register maintained within the governance support functionality.

In a multi-operator environment, the operators and other ecosystem participants could rely on a shared (potentially still distributed) policy registry similarly than they could use a shared service registry. Such a shared policy registry would essentially be a machine readable implementation of an ecosystem governance rulebook. Even if such a shared policy registry exists, the operators would still need internal governance support functionality that translates the shared policies to be implemented by the operator functionalities.

Uniquely in the scope of this element:

The governance policies' discovery and management are unique for the governance support functional element. The internal governance structures of the operator entity, such as a data ethics board and contractual settings for the leadership, can also be considered to be in the scope of the governance support functional element, even if their implementation is organisational rather than technical.

Key questions assessing the governance support capability of an operator are:

- What types of governance policies does the operator support?
- How does the operator maintain the governance policies up-to-date?
- How does the operator coordinate the implementation of the policies i.e. ensure that the policies are actually followed in its operations?

Scales of Interoperability:
Interoperability for governance support is defined by the extent to which aims and values shared between organisations are effectively translated into functional implementations.

Delivered with other functional elements:

- **Permission management** functionality uses the governance policies, clauses and contract templates from the governance support to generate compliant permissions and data sharing agreements.
- Enforcement of the governance policies from the governance support is delivered with corresponding functional elements. For example, applying governance policies for onboarding and expulsion of data sources and data using services happens via the **service management** functional element and required authentication standards are implemented by **identity management** functionality.
- **Logging and audit** functionality is responsible for the auditability of compliance with the governance policies.

Accountability and logging

Logging and accountability (LA) *functional element enables the maintenance of records (including record deletion) on data exchanges taking place and creating transparency about who accessed what and when and based on what permissions.*

Transparency and accountability are essential principles and prerequisites in many legislations. Accountability can enhance assurance, and logging can mitigate misuse or unintended use risks. Logging is not the sole responsibility of the operators and has counterparts in data sources and data using services.

Accountability arrangements may flow from the rules and regulations in the underlying governance framework, but many operators work without an explicit governance framework. Even in those cases, operators must comply with the relevant legislation that often includes logging and accountability obligations.

In general, governance implies some accounting obligations; but logging and accountability are still needed for auditability and transparency if no explicit governance applies.

Uniquely in the scope of this element:
This functional element covers logging and record management for internal monitoring, oversight and reporting (for authorities) and providing the individual with meaningful records and transparency of the activities.

Implementations of logging and accountability may provide:

- **Event and transaction logs:** audit logs, internal and external access logs, permission logs, and environment logging.
- **Characteristics of logging:** immutability, revocability, standard timestamping, and persistent history (independent of the operator that logged it).
- **Records for the individual:** metadata on the usage of data and services (what, when, how often etc.), and tracking of copies of data in circulation.
- **Log management:** configurable logging, log access, historical log management, and data minimisation of logs.

Key questions assessing an operator's accountability and logging capability are:

- What is logged and recorded?
- Who can configure logging and manage logs, and how?
- How is data minimisation and deleting records taken care of?

Scales of Interoperability:
Interoperability of the logging and accountability is defined by how accessible and portable the logs and records are for the other data ecosystem participants.

Delivered with other functional elements:
The actual logging implementation can be delivered with corresponding functional elements. For example, identity management can implement logging of identifiers and authentication events, and service management can implement logging of service registration and modifications.

3.2. Minimum interoperability requirements

The MyData community holds strong expectations for operators to cooperate and work towards interoperability. Actors with functionalities similar to those described above are not considered MyData operators if they do not embrace the vision of future interoperability between operators. MyData operators should be proactive on the journey to interoperability to allow for ecosystem growth and share resources.

There are many dimensions of interoperability, and clarity on specific objectives is required to progress our journey. We need to understand both the means for achieving interoperability and our ambitions for it. In the context of an interoperable MyData operator network, we identify four areas of focus:

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 within and between data spaces 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 that are entirely interchangeable with indistinguishable inputs and outcomes.

Interoperability provides overall system benefits at different, distinct dimensions that can and should be developed concurrently: technical (connectivity), semantic (informational), and organisational (governance, business models etc.) (Tolk, 2010).

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 useful for MyData. These could be data models for data control and governance (e.g., 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.

While we will work with other organisations to address opportunities for legal interoperability, for example, in the European Interoperability Framework (European Commission, 2017), it is currently beyond the scope of this paper and future work.

Organisational, semantic, and technical interoperability are all essential if we want to achieve ecosystems with multiple operators, data sources, and data using services that can work together to deliver human-centric services. Interoperability between the different actors in different roles is required to enable effective data flows in the ecosystem. People should not be locked into services but should be able to choose to move when they want to. The ability of the person to change their operator without barriers, or to use multiple operators, further requires that there is interoperability between operators.

By understanding the ecosystem roles and using the reference model for architecting the implementations, we can reach a degree of technical modularity that enables the separation of concerns (SoC). Each module addresses a different aspect, or concern, of the overarching system in this approach. When concerns are well separated, there are more opportunities for transparency and good governance.

The MyData community is uniquely placed on developing and driving frameworks for interoperable human-centric data sharing. We have both the skills and the mindset to ensure that interoperability questions related to personal data are correctly framed around the person's needs rather than the organisations' that should be serving that person.

Delivering human-centric interoperability requires agreement, alignment, and significant effort beyond just drafting rules or technical specifications. This journey towards convergence can be guided by an evolving roadmap where the immediate steps can be easily seen already, and further plans can be made as the situation unfolds.

Common goal: The first step of agreeing on a common goal has already been taken as our objectives are defined in the MyData declaration.

Common understanding and definitions: This next stage is embodied in this paper and the MyData operator awards process. They have created 'a state of common understanding' by defining agreed terms describing systems and methods.

Common processes: The output of this descriptive stage then allows us to identify common existing processes and common tasks.

Harmonising processes: We can then agree on which tasks in which functional elements of the reference model are the best targets for initial harmonisation efforts. Selection criteria may be their linkage (or lack of linkage) to other elements, their impact on the overall functionality of the ecosystems, or the ease or difficulty of the harmonisation task. Ultimately, however, the selection will come down to people and organisations wanting to take on any specific task.

Common governance: In parallel, we need to agree upon the position of MyData with respect to governance frameworks.

This early roadmap follows an action standard approach, where compliance is defined by completing the specified steps rather than being a quality standard.

The initial minimum interoperability requirement for a potential operator to be considered a MyData operator is to **describe the systems for personal data management with respect to the MyData operator reference model**. Operators need to show the modularity of their approaches as required by SoC. There will be aspects of an operator's service that are proprietary and other aspects that can contribute to best practice for open standardisation. The functions of proprietary service components must be described and the operation of non-proprietary components must be transparent. The interfaces between modules should be described in detail. This allows the community to identify the basic tasks commonly performed by most operators and build interoperable components from there.

This phase of describing operator systems based on the common reference model and terminology has encouraged the open sharing of practices and processes that have a common aim. The learnings from this approach will continue to inform the development of the roadmap towards interoperability described above and, ultimately, the emergence of rulebooks, auditable specifications, quality standards, and test tools. Mutual interoperability is inherently supported by this iterative way of working and the shared knowledge will help operators to innovate faster, better, and with lower risks to privacy.

3.3. Governance of human-centric data sharing ecosystems

Governance should be targeted at facilitating trust and opening up the ecosystems for innovation. Individuals should be protected, empowered to benefit from the data that organisations hold about them, and endowed with control over and visibility of how the data about them is used.

The ecosystem created by operators, working with data sources and data using services, is always part of a broader, social and economic system of individuals, communities, public organisations and private companies. Therefore, the ecosystem functions within the wider context of legislation, regulation, and social norms. Legislation is necessary for the creation of trust, but it is not sufficient. In order to create a level playing field in the market, rules of engagement between the different roles and actors fulfilling those roles are needed. This is often captured in an ecosystem governance framework (also called trust framework (Makaay et al., 2017)) which is binding at the ecosystem level.

Whether legal jurisdiction provides enough protection for an individual or not, governance codifies the explicit formulation of the re-balancing power that individuals are provided with by an operator. The level of an operator's responsibility towards the individual depends on the ecosystem. For example, in some ecosystems there is no strong governance structure in place, so a MyData operator has a correspondingly bigger responsibility of setting and enforcing the human-centric rules. As the MyData principles are independent of legal jurisdiction and the specifics of an ecosystem, they provide a universal guide to the setting of such human-centric rules.

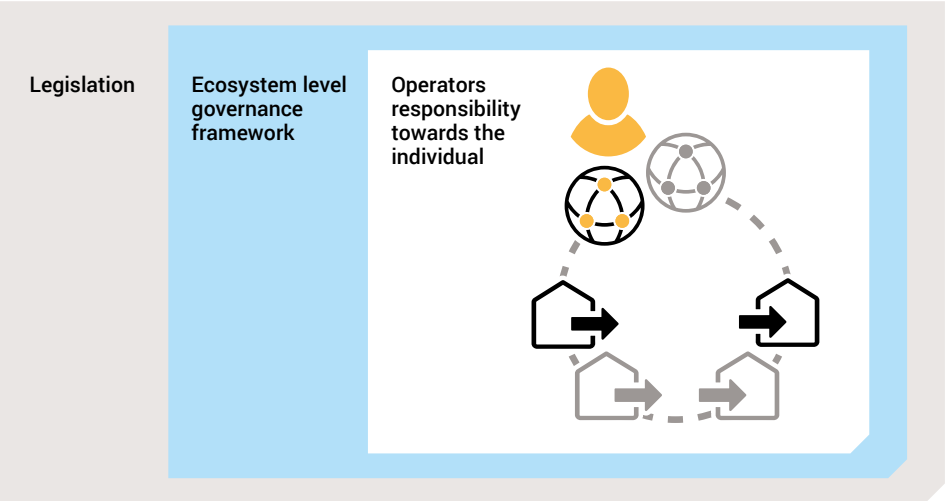

Figure 4: The tiers of governance in human-centric data sharing.

Legislation and soft law governing personal data ecosystems

In the European Union, the GDPR provides the authoritative basis for data exchange and privacy protection. Similar laws have been introduced in many places, including California, China, Chile, Japan, Brazil, South Korea, Argentina and Kenya (European Commission, 2019). Besides general privacy and data protection laws, various sector-specific regulations also govern data exchange, especially in health and financial sectors in many jurisdictions. Also, laws covering international data transfers and cyber security are relevant for data ecosystems. In many domains of data related regulation, the EU is moving first. Based on the EU Data Strategy (European Commission, 2020), the European Commission is rolling out many new laws that bring interoperability of data processing services and governance of so-called data intermediaries subject to regulation.

Governments can also incentivise personal data ecosystems via soft law initiatives such as the Japanese government initiated certification programme for Trusted Personal Data Management Services (Onga, 2019), and funding schemes such as the Korean government-led MyData programme for financial institutes (Minkyung, 2022)

Ecosystem governance frameworks

The governance of mature ecosystems is typically based on rulebooks (e.g., Sitra, 2019) and underlying contractually enforceable agreements between parties in the ecosystem. Such a governance framework describes the binding, ecosystem-wide rules and specifications (business, legal, technical, social) and defines the ecosystem's borders. Such governance frameworks may specify sanctions, auditing, and enforcement mechanisms for the rules. They can also help regulate data standardisation, validate data sources, manage permissions, enhance data portability, and establish ways in which individuals may influence the governance structure itself. Well-known examples in other domains beyond personal data management include credit card systems such as Visa, domain name registration systems governed by ICANN, or telecommunication frameworks governed by GSMA and the ITU-T. In the area of personal data management, Qiy (Netherlands) and HAT (UK) are early examples of governance frameworks.

An operator's responsibility towards the individual

An operator is always an infrastructure provider and an enabler for all participants of the ecosystem within which it operates. Operators have a duty to care for the individual's bespoke interest and should facilitate a more balanced and fair relationship between people and organisations. The degree of responsibility that an operator holds towards the individual will vary depending on the functionality they deliver, the strength of the applicable personal data legislation, and the maturity and kind of ecosystem governance in place.

We consider it most likely that there will ultimately be different degrees of operator responsibility. In every case, operators will need to determine the appropriate degree of responsibility towards the individual, balancing the strength of ecosystem governance and applicable regulations. It is illustrative here to consider the two extreme scenarios possible for degrees of responsibility: an operator with strong responsibility on the one hand and a more neutral operator on the other.

In the first scenario, there are cases where an operator is serving the individual with a very high level of responsibility. One approach is for operators to assume a fiduciary duty where, as fiduciaries, they must always put the person's interests ahead of their own (Balkin, 2016). Full fiduciary responsibilities restrict the choice of business models and may need to be backed by regulation (as seen with doctors and lawyers) to maintain a level playing field. Approaches with a voluntary, near-fiduciary degree of operator responsibility may be relevant to guarantee human-centricity in weakly governed ecosystems with little externally enforced regulation or commonly agreed rules to protect individuals' rights and interests.

At the other end of the spectrum, the operator is a more neutral actor with a lower degree of responsibility in setting or enforcing rules to guarantee human-centricity. This approach is relevant when strong ecosystem governance, strict regulation, or an effective combination of the two is in place. The legislator or the governance body then becomes the ultimate guarantor of human-centricity, and operators must follow the rules and regulations. The shared systems of governance and regulation increase confidence for the person while simultaneously reducing risks for operators and thus reducing costs and stimulating innovation.

There is a general understanding among the current operators that it is easier to start by having operators develop separate use cases. In such situations, the operators should hold strong responsibility towards the individuals. The minimum requirement at this stage for the MyData operators is to **demonstrate alignment with the MyData principles**. The development seems to be towards governed ecosystems and thus more neutral operators in the future.

Who controls the operator?

The control of an operator is a fundamental question when assessing how the principles of the MyData declaration will be applied and embodied. In our investigation, we identified five broad categories of entities that might run an operator. These categories are based on the relationships an individual could have with an operator. These categories are not mutually exclusive, and organisations may fall into more than one category depending on their legal status.

Business to person: Individuals are customers of the operators. For example, existing critical infrastructure operators such as banks, telecom operators, or utility companies could extend their services and become MyData operators. Also, new companies can be established based on this same commercial customer relationship.

Business to business: The individual's relationship with an operator is indirect and mediated by another service. For example, permission management functionality may be embedded in an end-user service that relies on an outsourced operator to provide that functionality.

Individual: Individuals themselves take responsibility for operating the infrastructure to interact with the rest of the ecosystem. This can happen, for example, by running their own personal data store (PDS) instance.

Collective: Individuals collectively support and manage an operator as members through the legal forms of associations, cooperatives, or data trusts. For example, existing patient associations, labour unions, or cooperative model companies could provide operator services to their members. Further, purpose-built data trusts and cooperatives are being experimented with in several places and domains.

Public: Individuals have a citizen relationship with an operator run by public authorities. For example, cities or other public entities could provide operator services, especially to facilitate the flow of personal data in public services.

Operators falling into different categories subsequently have different requirements for financial and human capital investments. They also have different risk profiles across areas such as financial sustainability, privacy, and centralisation. However, it will be possible for MyData operators to be created and managed in all of them.

The European landscape on governing data ecosystems

The EU Data Strategy (2020) sets the direction for developing and incentivising governed data ecosystems in Europe, the so-called **common European data spaces**. Among the initiatives based on the data strategy, particularly important for governed ecosystems and MyData operators, are **Data Governance Act (DGA), Data Act (DA)**, and establishing a framework for a **European Digital Identity (updating the eIDAS regulation from 2014)**[6]. These initiatives are logically connected from the perspective of MyData operators. However, it is worth noting that these were developed parallel to each other, and at the time of this publication, it is still early to say if and how they will be aligned. MyData operators may have a significant role in turning these regulatory initiatives into functional and human-centric data infrastructure. This landscape is in an early stage of maturity, and the coming few years will be crucial for its development.

From a MyData operator's perspective, a desirable scenario would be a harmonised set of regulations rooted in the principles of the GDPR, with sufficient support to initiate interoperable data spaces encompassing personal and non-personal. One way to imagine this scenario, specifically from the point of view of personal data, is as follows.

The GDPR sets the baseline for the allowed practices in data spaces. The Data Act builds on this foundation by suggesting and mandating measures to improve data re-use. It does this by setting interoperability requirements for ecosystem actors (especially what it calls "operators of data spaces") and strengthening data portability rights for data subjects and organisations.

The DGA, on the other hand, provides for the infrastructure to facilitate such re-use by establishing a new class of actors in data ecosystems, namely data intermediaries. MyData operators in the EU form a subset of these intermediaries. These intermediaries facilitate data-sharing relationships between ecosystem participants and across ecosystems. They can also – individually or as networks – function as "operators of data spaces" as described in the Data Act.

Finally, the potential of the eIDAS revision is to provide the keys to specific data spaces. Reliable identification of oneself (a natural or legal person) to others in the ecosystem or data space unlocks access to shared and accessible data one has permissions to process. Because different kinds of data have different requirements for the kind and level of authentication required to process it, an ideal outcome of the eIDAS revision would allow for multiple types of European Digital Identity Wallets fit for different purposes. A certain set of attributes, for example, would unlock access to a certain type of data or a specific data space. MyData operators, or data intermediation services in general, could potentially serve as issuers of such wallets.

6 Other significant pieces of legislation in their own right are the data strategy's Digital Markets Act (DMA), the Digital Services Act (DSA), and the AI Act, which however are less relevant for MyData operators.

Table 2: European regulatory and other initiatives relevant for MyData operators.

Data governance act (DGA) The DGA governs data intermediaries that offer services in the EU. The law also establishes a European Data Innovation Board that will propose interoperability guidelines for the data intermediaries and the European data spaces. The DGA enters into force during the first half of 2022. From there starts a 15-month transition period before the application of the law starts by the end of 2023. The DGA defines a 'data intermediation service' as a: *"service which aims to establish commercial relationships for the purpose of data sharing between an undetermined number of data subjects and data holders, on the one hand, and data users on the other hand, through technical, legal or other means, including for the exercise of data subjects' rights in relation to personal data"*. Providers of a data intermediation service will need to comply with the requirements of the DGA and submit a notification to a national authority in the EU. Article 11 lays out fifteen required conditions for providing data intermediation services. The interoperability requirement is essential for the MyData operators. It reads: *"the provider shall take appropriate measures to ensure interoperability with other data intermediation services, among others, by means of commonly-used open standards in the sector in which the data intermediation service providers operate"*. Registered providers of data intermediation services that comply with all requirements can operate in all EU member states. They may use the title 'provider of data intermediation services recognised in the Union' and a common logo to be issued by the European Commission.	MyData operators offering services in the EU are in the scope of the DGA and would therefore need to comply with its requirements.
Data act The Data Act proposal published in February 2022 (finalised in 2023) is a progressive legislative proposal to increase access to data for the users of connected products such as IoT devices and related services. The Act also covers interoperability related to data spaces and minimum requirements to smart contracts for data sharing, easier switching between cloud service providers, and business-to-government data sharing, among other topics.	The Data Act gives users of connected devices the right to share data with third parties. The data holders must also fulfil the data requests coming via DGA notified data intermediaries, such as MyData operators, acting on behalf of the user.
European digital identity (eIDAS 2.0) Proposed revision to the current EU regulation of electronic identification with an ambitious timeline to be in force by June 2024. This represents a move towards a user-centric identity model and the creation of European Digital Identity Wallets that would enable citizens' control over their data in identification and authentication processes.	Some MyData operators may become eIDAS wallet operators and the regulation is relevant to the implementation of the identity management functional element of MyData operator reference model.
Data spaces Data space is a decentralised infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles (Nagel and Lycklama, 2021). The European Commission references '*common European data spaces*' in the DGA, Data Act and the upcoming European Health Data Space regulation. The commission also dedicates funding for creating such data spaces.	Data intermediaries, such as MyData Operators, can provide infrastructure for the data spaces and have a specific role in connecting and creating interoperability across several data spaces.

3.4. Operator business models

No operator can be sustainable in the long run without a solid business model, whatever their legal status or type of control. Operators can be run commercially, as nonprofits or NGOs and public institutions.

In the long run, if true interoperability between operators is expected, there needs to be some convergence on business models to be compatible at the ecosystem level. Taking the example of telecom operators, they all work with the same basic business logic that the one who makes the call pays for the call. If different operators had different value capture mechanisms (say, one charged the caller, another charged the receiver, and the third added advertisements before the call and charged the advertiser), then the interoperability needed to roam between networks would have been much more challenging to achieve.

Again taking telecom operators as an example, we have seen that the breakup of the national telecom operator monopolies has resulted in a significant drop in call charges. Likewise, personal data ecosystems must provide individuals and organisations with options for mutual engagement that are superior to platform-based monopolies in terms of convenience and cost as well as privacy and ethics. The MyData operators and ecosystem participants must find similar opportunities to establish alternatives to monopolies that significantly remodel the cost and income structures of the incumbent market platforms.

Business models often evolve when the solutions become more mature. Subsidies, government funds or investments might be needed to create the ecosystem and all technical components or even public awareness of the solution. In the mature stages of the solution, the need for subsidising income decreases. However, some business models will stay dependent on government funds due to their societal nature (e.g., the government pays for roads and other infrastructure). The current business models of the operators we studied are not always clear and the sustainability of some models may be limited. This lack of clarity and limited sustainability are characteristics typical of a market that is yet to develop, where an ecosystem is still in the process of inventing itself. Many operators have already advanced beyond the initial pilot phase, but the scale on which they are used is often still limited, of course with some exceptions. Additionally, interoperability between operators is just emerging as a priority for the current operators. So far, bilateral agreements between operators, data sources, and data using services have been the norm.

There are costs associated with running an operator, such as providing compliance and security (including availability, utility, integrity, authenticity, confidentiality, nonrepudiation). Studying the business models of existing operators, we observe three revenue sources that cover these costs: **(1) revenue directly generated within the ecosystem** from data sources, data using services or the person, **(2) revenue generated from a subsidising customer** from outside the ecosystem, or **(3) the operator function supported by entirely different activities**. Currently, the first model is in its infancy, and many operators rely on the revenue from outside the ecosystem or subsidise the operator activity by other means.

As the operator market matures, more operators should move from the second and third models towards greater financial self-sustainability and revenue generated within the ecosystem. It is desirable because it removes commercial influence from outside the ecosystem and ensures that actors are not reliant on, for example, government subsidies. In the first model of revenue created within the ecosystem, there are several options for operators in terms of from whom and for what to charge.

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.

An operator may also need to share revenue with these actors or utilise other value exchange methods. Business models will balance these revenue streams against the costs of delivering services. It is important to recognise the separation between the fees for the data itself and the fees for the connectivity enabling data flows. The different fees may be combined at the point of billing, but for the sake of transparency and to maintain separation of concerns, they must be unbundled in the business model and its communication.

It is important to consider that some MyData operator businesses should become profitable in time. The operators' different control and governance models will result in differences in how the revenue is shared. We will need to judge if some control structures can be seen as more or less aligned with the MyData principles than others, but this remains future work.

In summary, there are a variety of operator business models currently in use and available in the future as the field matures. In terms of business models, the minimum requirement for the MyData operators is to show that they follow the two criteria of **transparency** and **the person as a primary beneficiary**. Information about the revenue flows must be as visible to the individual as the data flows, and where profits are made, they must be declared. We also recognise that individual agency in a market context requires the ability to pay and to be paid. However, we believe that we should consider the agency of people to extend well beyond the confines of the market. This is why a MyData operator will need to prioritise their duty of care for an individual over encouragement to monetise or overly share personal data. For example, a business model that emphasises the volume of data transactions might become unable to exercise their duty of care towards the person in cases where those transactions are not to the benefit of the person. As a result, we assert that the markets in which MyData operators exist should be markets for services rather than markets for data.

4. Future work

The objective of the first edition of this paper, published in 2020, was to create a common understanding of the functionalities and responsibilities of MyData operators and start a journey towards interoperability. The high-level descriptions of the most important functional elements that characterise an operator have allowed many operators to self-describe their offerings consistently, deepening our understanding of the functional aspects of trusted intermediaries. The development of initial minimum criteria for operators to be considered MyData operators created a platform for collective work focussed on building an interoperable network of operators.

In this second edition of the paper, we have furthered the descriptions of the reference model to help guide operators and establish some of the common technologies in use. The fundamental aim is to make the operation of infrastructures for personal data use easier for people and more human-centric in general. Our work to advance on the journey of interoperability has immediate benefits for individuals as interfaces, processes, and communications become standardised - reducing the effort required to adopt new services.

MyData operator reference model

We will continue to develop the depth and breadth of the reference model described in this paper and define requirements across the different dimensions of interoperability (technical, semantic and organisational). We will identify aspects of interoperability that are most reachable, set goals and create roadmaps for interoperability by functional element.

Mandatory requirements: as our work has developed, we have identified essential aspects of an operator's functionality. These include *identity management*, *personal data transfer and logging & accountability*. We will formalise the description of these requirements to create the first steps towards harmonisation.

Schema developments: in many functional areas *(permission management*, *service management*, *personal data storage*, *etc.)*, common schemas provide the best route to improving interoperability between operators and within data ecosystems. We will guide the adoption and creation of shared data models and semantics among operators to provide harmonised information exchange and communication. We will use commonly accepted standards, ontologies, libraries, or schemas available and support original works as necessary.

Thought leadership: MyData Global has proven its ability to convene thought leaders over recent years. We will use this platform to advance wider discussions about value exchange and *governance support*. We will also investigate how the services of MyData operators can be made more visible and accessible with technical service registries and so create a template for *service management* interoperability. We will continue to adapt and advance the reference model itself to better describe the evolving technical environment, for example, by addressing interoperability questions across different identity management paradigms and data verification.

MyData operator award

We will use the MyData operator award as the platform to lead our community engagement and wider communications. The objective is to demonstrate the need for interoperability for all stakeholders and accelerate operator-operator interoperability. The award will further develop the landscape of technologies, specifications and standards in use.

Award development: we will develop the MyData operator award to include robust, normative criteria in the mandatory functional elements of the reference model. We will recognise the maturity of the different service offerings and align award levels to regulatory requirements (specifically, the Data Governance Act). We will continue to advance common minimum standards where appropriate and promote standardised, publicly documented APIs.

Promotion & publications: we will publish more frequent, shorter updates on our progress, including our 2022/23 plans, analysis of the 2022 award submissions, reference model updates and application guidance to support the 2023 MyData operators award. We will build case studies of successful collaborative work between operators and document the already prevalent interoperability in the existing MyData operators. We will develop new visualisations of the technical relationships between functional elements and the organisational relationships between ecosystem participants.

Data ecosystems

During the last few years, we have seen significant development in new data ecosystems. Public agencies and governments facilitate some of these ecosystems, and such government-led ecosystems may even be defined in law, as is the case with the European Health Data Space. However, most are formalised in one way or another by collaboration agreements (rulebooks) between companies and other independent organisations. Some are called data spaces, and some have other names. MyData operators and the operator reference model will be developed to align with the intermediary roles in these ecosystems. A MyData operator functions as an intermediary in multiple personal data ecosystems and facilitates collaboration between ecosystems. Ultimately, this will build up a stack of MyData operator compatible ecosystems. There will also be less formal ecosystems and ecosystems created by a single operator. We will develop the MyData operator governance and overall functional framework to accommodate multiple and varying types of personal data ecosystems while retaining overall architectural coherence and the core principles of MyData.

Conclusions

The idea of human-centric personal data is gaining widespread traction globally. The MyData declaration defines the role of the MyData operator and sets out high-level principles for a human-centric approach to personal data. In personal data ecosystems, the infrastructure operators are in key positions to implement these principles and make human-centricity work in practice.

The MyData vision highlights competition in an open ecosystem where multiple providers of infrastructure-level services are mutually interoperable and substitutable. We use the metaphor of the 'journey of interoperability' for the work needed to progress towards such a global network of many competing and mutually interoperable operators. To initiate this journey, MyData Global used its 'power to convene' to bring together organisations that today run and develop operator-like services and related products and technologies. This paper is the studied result of the interactions with these operators.

As a 'state of the common understanding' among the operators, this paper presents the MyData operator reference model. It initiates discussions on operator interoperability, the governance of human-centric data sharing, and the business models available for operators.

The reference model lays out nine core functional elements an operator may have: (1) identity management, (2) permission management, (3) service management, (4) value exchange, (5) data model management, (6) personal data transfer, (7) personal data storage, (8) governance support, and (9) logging and accountability.

Interoperability between operators should be framed in terms of the needs of the person rather than the organisations in a given ecosystem. After acknowledging this as our goal and describing some common tasks and our approaches to minimum interoperability requirements, more robust requirements will be co-developed based on the reference model.

Governance of human-centric data sharing can be conceptualised at different levels. Legislation is the widest and least specific level. Ecosystem-level governance frameworks set more specific rules for the participants of a given ecosystem. Finally, the operators will have certain responsibilities towards the individual. The responsibilities of an operator will vary depending on the strength of the ecosystem governance and the regulation.

Operator business models should be transparent and designed with individuals as primary beneficiaries.

The results of this paper reflect a substantial advance in thinking on the topic introduced as 'trusted intermediaries' and described throughout as MyData operators. While the outcomes have been advanced considerably in this second edition, we recognise that follow-up collaborations are needed to iterate, evolve, and make them even more helpful. We hope that this paper will stand the test of time as the foundational basis for co-developing the idea and implementations of MyData operators and guiding the journey of interoperability. At the same time, some aspects of this paper may soon become outdated as the growing community of operators and other actors in personal data ecosystems progress on the issues laid out in the future work section.

We invite you to contact us if you would like to comment on this paper, learn more or join our community:

- Contact: operators@mydata.org
- Operators page: https://mydata.org/operators

Glossary

Term	Definition
Actor	An organisation or an individual performing one or more *roles*.
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 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 sharing agreement	All parties of the data transaction agree and comply with data sharing agreements that set out the purpose of the data sharing, cover what happens to the data at each stage, set standards, and help all the parties involved in sharing to be clear about their roles and responsibilities.
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.
Data using service	The *role* responsible for processing personal data from one or more d*ata sources* to deliver a service.
Distributed Ledger Technology (DLT)	A distributed ledger (also called a shared ledger or distributed ledger technology or DLT) is a consensus of replicated, shared, and synchronised digital data geographically spread across multiple sites, countries, or institutions. Unlike with a centralised database, there is no central administrator.
Ecosystem	The overall system created by the activities and connections of a set of *actors* and infrastructure interacting according to a common set of rules. Multiple ecosystems can exist, overlap, and collaborate.
Governance	A system of rules, practices, and processes used to direct and manage an *ecosystem*. The four pillars of good governance are transparency, fairness, accountability, and security.
Immutable logging	An immutable audit log is a tamper-resistant recording of how a system has been used.
Individual	A natural, living human being.
Interoperability	The ability of different systems to work in conjunction with each other and for devices, applications or products to connect and communicate in a coordinated way, without effort from the person. In this paper we use the Levels of Conceptual Interoperability Model (Tolk, 2010) with high-level classifications of technical, semantic and organisational interoperability.
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*.
Operator network	A group of *operators* with some degree of mutual *interoperability*.
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 roles.
Policy register	A policy register technically maintains uniquely referenceable versions of the governance policies and serves these policies in machine readable format.
Proto-operator	A product, service, or organisation that is in one way or another performing the *role* of an *operator* in personal data *ecosystems* or offers related tools, services, or technologies. Proto-operators come in many forms and under many different names and may cover one of more functional elements in the MyData operator reference model. They constitute the first generation of real-world MyData operators.
Role	A function or set of responsibilities for a particular purpose.
Semantic data model	Semantic data model refers to a data model describing the actual substance data (content). See also 'transaction data model'.

Separation of concerns (SoC)	A principle by which a modular approach to the development of a system is adopted. This approach entails each section addressing a different aspect (concern) of the overarching system. In the context of SoC in the personal data *ecosystem*, processing, storing, aggregating, displaying, governing data are concerns that need to be managed in a modular, transparent manner. SoC enables more opportunities for module upgrade, reuse, and independent development.
Self-sovereign identity (SSI)	An approach to digital identity that gives individuals control of their digital identities. SSI addresses the difficulty of establishing trust in an interaction. To be trusted, one party in an interaction will present credentials to the other parties. Those relying parties can verify that the credentials came from an issuer they trust. In this way, the verifier's trust in the issuer is transferred to the credential holder. This basic structure of SSI with three participants is sometimes called "the trust triangle. For an identity system to be self-sovereign, users control the verifiable credentials that they hold and their consent to use those credentials. In an SSI system, holders generate and control unique identifiers called decentralised identifiers. Most SSI systems are decentralised, where the credentials are managed using crypto wallets and verified using public-key cryptography anchored on a distributed ledger.
Transaction data model	Transaction data models enable the participants of the data ecosystems to share and manage data. Transaction data models define, for example, the structure of identity claims, permissions, service definitions, standard APIs, governance policies and log data syntactics.

References

Abiteboul, S., André, B. and Kaplan, D. (2015) '*Managing your digital life with a Personal information management system*', Communications of the ACM. New York, NY, USA: ACM, 58(5), pp. 32–35. Available at: https://dl.acm.org/doi/10.1145/2670528 (Accessed: 22 April 2020).

Balkin, J. M. (2016) '*Information Fiduciaries and the First Amendment*'. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2675270 (Accessed: 14 April 2020).

Ctrl-Shift (2014) *Personal Information Management Services: An analysis of an emerging market*. Ctrl-Shift. Available at: http://www.nesta.org.uk/publications/personal-information-management-services-analysis-emerging-market (Accessed: 14 April 2020).

European Commission (2016) '*An emerging offer of 'personal information management services'* – Current state of service offers and challenges, Digital Single Market'. Available at: https://ec.europa.eu/digital-single-market/en/news/emerging-offer-personal-information-management-services-current-state-service-offers-and (Accessed: 19 April 2020).

European Commission (2017) '*European Interoperability Framework – Implementation Strategy*', EC COM(2017) 134 final. Available at: https://ec.europa.eu/isa2/eif_en (Accessed: 22 April 2020).

European Commission (2019) '*General Data Protection Regulation: one year on, European Commission*'. Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_19_2610 (Accessed: 14 April 2020).

European Commission (2020) '*European Data Strategy*'. European Commission. Available at: https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf.

European Commission (2021) '*Final compromise text (Dec 10. 2021) of the EU Data Governance Act proposal*'. Available at: https://data.consilium.europa.eu/doc/document/ST-14606-2021-INIT/en/pdf

Gordijn, J. (2004) '*5 - e-Business value modelling using the e3-value ontology*', in Currie, W. L. (ed.) Value Creation from E-Business Models. Oxford: Butterworth-Heinemann, pp. 98–127.

Haaker, T., Faber, E. and Bouwman, H. (2006) '*Balancing customer and network value in business models for mobile services*'. Available at: https://doi.org/10.1504/IJMC.2006.010360 (Accessed: 21 April 2020).

Hafen, E. (2019) '*Personal Data Cooperatives – A New Data Governance Framework for Data Donations and Precision Health*', in Krutzinna, J. and Floridi, L. (eds) The Ethics of Medical Data Donation. Cham (CH): Springer. Available at: https://link.springer.com/chapter/10.1007/978-3-030-04363-6_9 (Accessed: 18 April 2020).

Hagel, J. and Singer, M. (1999) '*Unbundling the Corporation*', Harvard business review. Available at: https://hbr.org/1999/03/unbundling-the-corporation (Accessed: 14 April 2020).

Janssen, W. et al. (2019) '*Discussion Paper What is the MyData Operator?*', MyData Global. Available at: https://mydata.org/wp-content/uploads/sites/5/2019/09/Discussion-paper-MyData-operator-final.pdf. (Accessed: 18 April 2020)

Karhu, K. et al. (2020) '*Proposal of minimum interoperability mechanism for personal data*', Open & agile smart cities OASC Minimum Interoperability Mechanism (MIM). Available at: https://oasc.atlassian.net/wiki/spaces/OASCMIM/pages/30179329/MIM4%2BPersonal%2BData (Accessed: 24 April 2020).

Kuppinger, M. (2012) '*Life Management Platforms: Control and Privacy for Personal Data*', KuppingerCole. Available at: https://www.kuppingercole.com/report/advisory-lifemanagementplatforms7060813412 (Accessed: 14 April 2020).

Lanier, J. and Weyl, G. (2018) '*A Blueprint for a Better Digital Society*', Harvard Business Review, 26 September. Available at: https://hbr.org/2018/09/a-blueprint-for-a-better-digital-society (Accessed: 11 July 2019).

Lehtiniemi, T. (2017) '*Personal Data Spaces: An Intervention in Surveillance Capitalism?*', Surveillance & Society, 15(5), pp. 626–639. Available at: https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/6424 (Accessed: 19 April 2020).

Makaay, E., Smedinghoff, T. and Thibeau, D. (2017) '*Trust Frameworks for Identity Systems*'. Available at: https://www.openidentityexchange.org/wp-content/uploads/2017/06/OIX-White-Paper_Trust-Frameworks-for-Identity-Systems_Final.pdf (Accessed: 24 April 2020).

MIC Japan (2018) '*Release of the Guidelines of Certification Schemes Concerning Functions of Information Trust ver. 1.0*', Ministry of Internal Affairs and Communication Japan. Available at: https://www.meti.go.jp/english/press/2018/0626_002.html (Accessed: 18 April 2020).

Min-kyung, J. (2021) '*Government-led consumer finance data service kicks off*', *Korea Herald*. Available at: http://www.koreaherald.com/view.php?ud=20211201000676 (Accessed: 10 March 2022).

MyData Global Network (2017) '*Declaration of MyData Principles*'. MyData Global Network (before the MyData Global association was established). Available at: https://mydata.org/declaration (Accessed: 14 April 2020).

MyData Global (2019) '*What Is the MyData Operator?*', Workshop at the MyData 2019 conference. Available at: https://mydata2019.org/programme-page/what-is-the-mydata-operator (Accessed: 14 April 2020).

MyData Global (2020) '*MyData Operators thematic group*'. Available at: https://mydata.org/groups/mydata-operators (Accessed: 14 April 2020).

Nagel L., Lycklama D. (2021) '*Design Principles for Data Spaces. Position Paper. Version 1.0.*'. Available at: http://doi.org/10.5281/zenodo.5105744 (Accessed: 10 March 2022).

Obar, J. A. and Oeldorf-Hirsch, A. (2018) '*The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services*'. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757465 (Accessed: 19 April 2020).

ODI (2018) '*Defining a 'data trust''*, Open Data Institute: The ODI. Available at: https://theodi.org/article/defining-a-data-trust (Accessed: 14 April 2020).

Poikola, A., Kuikkaniemi, K. and Honko, H. (2015) '*MyData – A Nordic Model for human-centered personal data management and processing.*' Ministry of Transport and Communications. Available at: http://urn.fi/URN:ISBN:978-952-243-455-5 (Accessed: 14 April 2020).

Project VRM (2008) '*Project VRM – Berkman Centre*', Harvard University. Available at: https://cyber.harvard.edu/projectvrm/Main_Page (Accessed: 20 April 2020).

Rikken, M., Janssen, W. and Duits, I. (2019) '*Het landschap van Persoonlijk Data- Management*', InnoValor. Available at: https://drive.google.com/file/d/1IZDHRkOzGGOn_CzZxQxG4dB3lk9KAGtj/view, https://innovalor.nl/digitale-wendbaarheid/persoonlijk-datamanagement (Accessed: 17 April 2020).

Sitra (2019) '*Rulebook for Fair Data Economy – Rulebook Template for Data Networks*', Sitra. Available at: https://www.sitra.fi/en/news/a-new-rule-book-sets-out-the-guidelines-for-a-fair-data-economy (Accessed: 14 April 2020).

Sitra (2020) '*IHAN Blueprint 2.5*', Sitra. Available at: https://www.sitra.fi/en/articles/ihan-blueprint (Accessed: 22 April 2020).

Tolk, A. (2010) '*Architecture constraints for Interoperability and composability in a smart grid*', Power and Energy Society General Meeting, 2010 IEEE. Available at: https://www.researchgate.net/publication/224178883_Architecture_constraints_for_Interoperability_and_composability_in_a_smart_grid (Accessed: 14 April 2020).

Wang, F. and De Filippi, P. (2020) '*Self-Sovereign Identity in a Globalized World: Credentials-Based Identity Systems as a Driver for Economic Inclusion*', Frontiers in Blockchain, 2, p. 28. Available at: https://www.frontiersin.org/articles/10.3389/fbloc.2019.00028/full (Accessed: 19 April 2020).

World Economic Forum (2013) '*Unlocking the Value of Personal Data*', World Economic Forum. Available at: http://www.weforum.org/reports/unlocking-value-personal-data-collection-usage (Accessed: 21 April 2020).

World Medical Association (2018) '*Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects*', World Medical Association. Available at: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects (Accessed: 17 April 2020).

Appendix 1 – Awarded MyData operators

OwnYourData Austria Awarded: 2020 2021 2022	OwnYourData is a non-profit association and helps you to achieve unrestricted access to your data for your benefit. We offer different services and products – all licensed as Open Source and according to the MyData principles: • Data Vault: Tap into data sources and store your information in a personal vault. Based on that data OwnYourData presents you correlations and insights. • Semantic Container: Enable secure and traceable data exchange between multiple parties. The solution is standards-based and offers a lightweight infrastructure to make open and commercial data available in an auditable and reproducible manner. Notary: This service offers the safest and easiest place to protect your data and documents. Based on blockchain and cryptographic technology, your information is anonymous, tamperproof and legally binding. • MyData Weekly Digest: Read about current news and events in the human-centric data ecosystem or browse apps and services you can use to manage personal information Website: https://www.ownyourdata.eu
Geens Belgium Awarded: 2020 2022	Geens is a TRUST platform for secure and private data sharing for individuals, teams, companies and communities dealing with private or sensitive information. Geens combines encrypted data storage, close-to-zero-knowledge security, blockchain timestamping, micro payments, and user engagement - all in one platform. Geens is a non-profit, non-governmental and non-corporate organisation which provides technology services to protect individuals. The Geens Ethical Committee governs the compliance with the rules of conduct, standards and policies that guide the Geens developments. This is to guarantee absolute independence and interoperability to strengthen the role of the trust provider. Geens, as a non-profit organization, can never be acquired by anyone. All profits return into the Geens ecosystem to benefit its members. Website: https://geens.com/
Meeco Belgium, Australia, United Kingdom Awarded: 2020 2021 2022	Pioneering since 2012, Meeco unlocks the power of consent-based and permissioned personal data, giving organisations the tools to empower their customers to access, control and create mutual value from their personal data. Meeco offers FinTech, RegTech, LifeTech and KidTech solutions across a range of sectors including banking, telecommunications, government, health and ESG. Meeco's technology is a Privacy and Security by Design, ISO certified suite of products: a decentralised wallet (incorporating identity, verifiable credentials along with micropayments and tokens), a key management store, a secure data enclave (vault) and a credential brokerage service decentralised on Hedera Hashgraph. Meeco's technology is a strategic enabler for organisations that wish to become a MyData Operator, offer data intermediary services, or provide GDPR (Europe), CDR (Australia), or CCPA (USA) solutions. Website: https://www.meeco.me
Datavillage Belgium Awarded: 2022	Datavillage's mission is to unlock the value of personal data. Datavillage has developed a privacy-preserving personal data platform that allows companies to access and process end-users behavioural data, in a legal and transparent way, without having to collect the data in their own systems. The platform serves several use cases of hyper-personalization in different industries as being used to improve recommender systems and offer end-users with a personalised experience using the data they produce on different platforms while preserving their privacy. This way, the trust between the service providers and the end-users is strengthened while offering added- value services that fit who the users truly are, always preserving their privacy rights and making them remain in control of the personal data all along its lifecycle. Website: https://www.datavillage.me

Diabetes Services Denmark Awarded: 2020 2021 2022	Diabetes Services is building a human-centric digital ecosystem of data using services for diabetes- and lifestyle management for better health and quality of life. Every day, digital sources like smartphones, medical devices and personal wearables generate a large amount of data of the person using these sources. We help gather this personal data to assemble a digital diary so that it can be converted into decision support for practical action and value-based diabetes treatment where focus is on the individual person. The ecosystem is the link between persons, healthcare, business and society which eases free movement of all real life data sources to foster opportunities for faster innovation, research, and sharing of personal data. We provide full IT services that all actors can use to contribute to interoperable services in an open ecosystem.We also provide a governance model for a joint development of IT solutions with microservices based on open source and standard components. Website: https://diabetes.services
DATA for GOOD Denmark Awarded: 2022	DATA for GOOD Foundation (DfG) is a Danish not-for-profit organisation that promotes awareness of individual data rights and equip citizens with digital tools that allow them to exercise their right to data portability (GDPR, art. 20) and thus make more data available for common good purposes. DfG is a neutral intermediary governing a personal infrastructure based on a secure computation technology (Multi Party Computation - MPC). Website: https://dataforgoodfoundation.com
Sensotrend Finland Awarded: 2022	Sensotrend makes life with diabetes easier. We combine data from medical devices and wellness trackers, derive actionable insights from the data, and facilitate sharing the data with healthcare professionals, peers, and for research purposes. Website: https://www.sensotrend.com
MyDataShare Finland Awarded: 2020 2021 2022	MyDataShare is a MyData operator platform and a MyData operator developed and provided by Vastuu Group Oy. Vastuu Group is a data and ICT service company developing public-private-people ecosystems. MyDataShare provides multiple perspectives into a holistic, integrable, and extensible platform product that covers commercial, legal, technological, and ethical perspectives of personal data use to assist customers to comply with legal and ethical requirements in implementing digital services. MyDataShare provides intuitively easy identity and permission management functionalities for customers and end-users. MyDataShare does not store personal data, it only brokers access to personal data. MyDataShare operatorship is licensed to customers as Software as a Service or as a customer-operated platform license. MyDataShare collaborates with Open & Agile Smart Cities on Minimal Interoperability Mechanism specifications and reference implementations. Website: https://mydatashare.com
Be Swarm France Awarded: 2022	Be Swarm offers a new milestone in personal data ownership. The paradigm of current management of users' data neeeds an overall disruptive approach. Our solution completely rethinks the management of data handled by applications we use on a daily basis by providing a global response to all current problems. We guarantee users that they are the only ones who can access any of their personal data. All personal datas handled by applications are gathered together and stored in a single secured place belonging exclusively to users. Therefore, users are able to switch apps without ever losing data again because apps no longer hold any of the data. Users can freely choose where their personal data are stored. Applications via Be Swam handle their personal data without any possibility of identifying users. Therefore, it is 100% anonymous with regards to applications. So, it becomes 100% safe to apps' users. At last, thanks to Be Swarm, apps' users regain full control of their personal data. Website: https://beswarm.fr/index_en/

Cozy France Awarded: 2020 2021 2022	Cozy Cloud provides users with a personal cloud, Cozy. French and open source, Cozy is a human-centric and privacy-respectful digital home, in which users collect all of their data expanding their digital possibilities with personalized services. Data are spread in silos on the Web and until now out of control of the most legitimate to own them: individuals. It aims at giving back to users their personal data and help them into managing it through data connectors and functional applications. Open source, Cozy can be examined by experts, is "forkable" when necessary and moveable in just a few clicks. That's exactly why users can trust Cozy: "You will stay because you can leave" is one of Cozy Cloud mantras. Cozy gives users the possibility of having a personal assistant that is not an advertising assistant like Alexa, Google Home or Amazon Echo. Cozy is a way of using data to emancipate, rather than manipulate users. A big change for the paradigm of data! Website: https://cozy.io/en/
Fair&Smart France Awarded: 2020 2021 2022	Fair&Smart is a French SaaS company founded in 2016. Our mission statement is to make personal data accessible and useful for the good of the people and the society. Our platform connects data sources, data using services and individuals in any life area to allow secure and consented personal data transfers with full auditability, privacy compliance and end-to-end encryption. Myfairdata is the name of the platform for individuals: a web and mobile application allowing people to store and share data, manage permissions and send GDPR requests (portability, access…) Right Data is the name of the platform for organisations: a web application and REST API allowing organisations to make personal data available for third parties and/or access to data made available by third parties under the control of the individuals, with GDPR compliance built-in and premium Privacy UX. Two white labelled modules based on the platform are also available : Right Consents (Consent Management Platform) and Right Requests (SRR automation tool). Those solutions are frequently distinguished by technology specialists like Gartner or Wavestone. Website: https://www.myfairdata.com/en
Onecub France Awarded: 2020 2021 2022	Onecub is a French-based MyData Operator allowing the creation of data-sharing ecosystems (or data spaces). It allows individuals to share their data between their digital services through a seamless user experience and full privacy control. Onecub works mostly in the Tourism/Mobility and Agri-food sectors. Our vision is to promote the creation of a Fair Data Economy relying on human-centric data spaces. Onecub approves the MyData Declaration and all its principles. We are co-founder of the aNewgovernance AISBL based in Brussels where we build a common operational model for the data spaces, with fellow data spaces builders in other sectors like skills, health, space data. Onecub's main project in 2021 is the creation of a major data space in France in the tourism & mobility sectors. Since Paris will welcome the Olympic Games in 2024, we will leverage the event to align public and private interests in order to create a flagship data space. Website: https://www.onecub.com
Visions France Awarded: 2020 2021 2022	Visions provides a human-centric data intermediary service allowing people and organisations to share data in a secure, ethical, legal and easy way. Visions focuses on skills data helping people in their personal and professional development. People can manage their authorisations from a central dashboard and organisations can easily request data from other services for specific purposes through Visions API. Visions is a data ecosystem orchestrator, managing legal, technical and business aspects of the data sharing. Website: https://visionspol.eu
polypoly Germany Awarded: 2020 2021 2022	The polyPod ensures that personal data is stored on the end device. It is no longer necessary to send data to central servers, although this is still possible. Alternatively, the data is evaluated decentrally on the end devices of the individual, who can simply download the polyPod to all their end devices like an app. For processing, the polyPod uses the unused computing power of the end devices to map the functionalities of a server. With the polyPod, individuals have a private server that corresponds across all their personal devices and that they can control at any time. Every individual can give third parties access to this server and thus to personal data and computing power. To other citizens, to entrepreneurs or public servants, against payment or as a donation. Who may access their polyPod or use the computing power, and for how long, is determined by each user at any time. Website: https://polypoly.coop

SOWL Germany Awarded: 2021 2022	esatus AG is a medium-sized consulting company, founded in 1999, with its headquarter in Langen, Germany. esatus AG's mission is "Enforcing Information Security", with strong expertise in Identity & Access Management (IAM). Within its proactive innovation management, esatus AG is engaged in blockchain technology since 2016, particularly in the area of identity management. We are focused on distributed ledger-based solutions for Self-Sovereign Identity (SSI) and now have applied SSI to the IAM domain. Since 2019, esatus AG merged its many years of experience and outstanding know-how in these fields as well as its software development proficiency into the SOWL solution. SOWL is a comprehensive suite for Enterprise IAM, fully embracing SSI principles and technologies. SOWL is dedicated to build bridges between classic IAM protocols and technologies and the new, future-proof world of SSI. It delivers a business-friendly technical implementation of tailor-made IAM concepts. Website: https://esatus.com/solutions/self-sovereign-identity/sowl/?lang=en
Trinity IDP Germany Awarded: 2020 2021 2022	Trinity IDP is a mobile white-label SDK that performs all the functions of an identity provider on the mobile device and combines the requirements of SSI and OIDC in one solution. Website: https://www.comuny.de
My Information Tracer Japan Awarded: 2021 222	My Information Tracer (aka mint) is the platform for distributing Personal Data in Japan. It takes role as a hub among enterprises, and provides basic functions necessary for Personal Data distribution. Website: https://www.nttdata.com/global/en
paspit Japan Awarded: 2020	paspit is a smartphone and web application with which its users can store their personal data and share it with others based on his/her consent. The personal data comes to paspit from mainly 4 types of data sources. • Manual data entry (name, email, gendar, occupation etc..) • Survey responses. (What are major life events in your life in recent years? Do you own a house? etc) • Personal data scraped from web services (Amazon, netflix, etc.). • Personal data collected via APIs (Google calendar). paspit operators invite businesses to join the platform and they play the role of data using services. Those data using services send data sharing requests which should have at least one benefit to users. Once the users accept the requests his/her personal data is shared with the requester who is obliged to give back the benefit. If the users change their mind, they can withdraw the consent. Website: https://paspit.com/
Personium Japan Awarded: 2020 2021 2022	Personium is the first and still the only open-source personal data store platform in Japan. The platform was created by FUJITSU Limited back in 2008. It is developed for both common individual and enterprise users. Personium provides the following features: • Empowers individual to manage permission of data access • Empowers individual to manage aggregated data • Supports data portability by using well-known standards • Supports transparency by offering open-source core components and apps • Supports interoperability in different technical levels Fujitsu Limited is the leading Japanese information and communication technology (ICT) company, offering a full range of technology products, solutions, and services. Website: https://personium.io

Ockto Netherlands Awarded: 2020 2021 2022	Ockto is a platform with which persons can collect data from different data sources and pass this data on to a data using service (provider). The solution ensures that persons can quickly and easily collect information and share it with their adviser, bank, mortgage lender or other service provider. The person collects the data desired by the data using service with the help of Ockto, and after reviewing that data, optionally transfers this data. Ockto charges a transaction fee from the data using service for this service. Ockto's strategy is aimed at achieving the following objectives: • Being a reliable party for consumers • Acquiring a sustainable position within various chains in the Netherlands • Being a reliable partner of data sources (government and banks) • Extending Ockto with additional data sources • Becoming a key cabinet for the consumer to his / her personal information. • Internationalization of Ockto to other European countries. Website: https://www.ockto.nl/
Financieel Paspoort Netherlands Awarded: 2022	Stichting Financieel Paspoort is a non-profit foundation that strives to improve the financial resilience of the individual. This is done by providing tools and developing standards that enable the individual to retrieve and share personal financial data easily. We motivate organisations to open up their databases and work together to develop and implement standards. In the meantime we create value by enabling individuals to make use of secure methods and techniques that are currently available and which the individual is entitled to use today. Personal financial information can also be shared to make the provision of support and advice more easy and more accesible. Connections with external advice services, for which personal financial data is required, can be established, relevant data can be shared automatically and added value can be created, all under control of the individual. Website: https://financieelpaspoort.nl
IRMA Netherlands Awarded: 2022	With IRMA you can manage your digital identity on your mobile phone. It is easy to log in and make yourself known, by disclosing only relevant attributes of yourself. For instance, in order to watch a certain movie online, you prove that you are older than 16, and nothing else. You can also sign documents digitally. You use only relevant attributes of yourselves in a digital stamp. In this way you can sign with IRMA as a medical doctor, or as citizen, or in some other role. Data in IRMA come from trusted sources and are cryptographically protected. Thus, the attributes that you disclose to make yourself known are genuine and are really about you. IRMA.app is a Privacy by Design, Decentralized and Open Source solution powered by SIDN.nl (= solid base) Website: https://irma.app
Schluss Netherlands Awarded: 2020 2021 2022	With Schluss, you – and you alone – decide who knows what about you. With Schluss you get a digital vault in which ALL your data is securely stored. From simple addresses to complex financial and medical records. You decide to whom you disclose your data, for what reason and for what period. You keep an overview over your data and any disclosures. So that you can keep control over them and act as a data operator yourself. Schluss knows nothing of its customers and doesn't even know when a new vault has been opened. The information you entrust organizations with is all up to date. They have access to real time up-to-date customer data, comply to GDPR and don't have to store this personal information themselves. And where the Schluss vault is as closed and secure as can be; the organization behind it is open and transparent. Schluss will be a worldwide cooperative, with users as co-owners. Schluss is a movement, representing all internet users. Together we can change the Internet! Website: https://schluss.org
EYD Norway Awarded: 2022	EYD delivers a platform for services that helps personal users and companies safely retrieve, secure and enable personal data in compliance with privacy regulations. By providing a simple, secure and trusthworthy plattform for services to retrieve and enable personal data, we aim to create valueable connections between your business and your users. By empowering your users with insights, we help strengthen the credibility of your brand and hence the value of your users. Website: https://eyd.tech

Fairdrop Slovenia Awarded: 2020	Fairdrop is an open source personal data storage solution aiming to achieve user sovereignty over and privacy for their data. Data can be encrypted locally in the user's browser before being stored or shared with other accounts with peer-to-peer Swarm storage as the back-end. The peer-to-peer nature means no setting up of infrastructure is needed. There is no central authority to rely on or to block access to the data. The amount of storage used can be scaled as needed. This covers use cases for individuals storing their data as well as apps storing individual specific data, and others. An example use case is the storage of Kantara compliant consent receipts in a special "folder". Faidrop is leveraging Fair data society open source javascript libraries for working with Swarm storage, blockchain and user accounts. Website: https://fairdrop.xyz/
MyDataMood Spain Awarded: 2022	MyDataMood is a platform for managing personal data and digital identity, which allows a transparent, fair and honest exchange of data between Citizens and Organizations. Website: https://mydatamood.com
VALENCIADATA Spain Awarded: 2020 2022	VALENCIADATA is a data operator for managing and using relevant personal information for research activities. Our aim is to empower the citizens to control their personal data. The citizens may give data access to different innovative agents such as research institutions. Our objective is to develop a digital platform that includes the gathering, classification and storing of personal data for research projects, as well as all the interfaces that the citizens, companies and other innovation agents need to access to the system. The platform will be compliant with the legal frame established in the European General Data Protection Regulation (GDPR). The main objectives are: • To facilitate the reuse of personal data for research if the individual provides informed consent. • To connect services that can use the personal data in order to provide useful information to the individuals. Website: https://valenciadata.ibv.org
iGrant.io Sweden Awarded: 2020 2021 2022	iGrant.io is a data exchange platform that helps organisations to access personal data in an ethical and sustainable manner. Using iGrant.io's data exchange services, organisations gain access to verifiable, auditable and data regulatory compliant personal data. Every data exchange has an associated Data Agreement (DA) that records conditions for an organization to process personal data in accordance with data regulations, such as the GDPR. The key value propositions offered by the iGrant.io platform are: • Improved access to high-quality personal data by providing transparency and empowering users to control data usage. • Compliance by design reducing the risk of non-compliance to data regulation when it comes to personal data usage, by linking Data Agreements to a legally endorsed Data Process Impact Assessment (DPIA) process. • End-user SDKs (e.g. Data Wallets, User preference centre) that can be embedded into existing mobile applications and portals. Website: https://igrant.io/
BitsaboutMe Switzerland Awarded: 2020 2021 2022	BitsaboutMe is an innovative online platform where users can securely manage their digital life and make fair data deals with companies and organizations. At the heart of BitsaboutMe is the privacy of each individual user. They can merge their online accounts in one place, get a transparent 360-degree overview of their digital lives and thus regain full control over their personal data. The marketplace function enables users to share parts of their personal data securely with companies and organizations for a reward or anonymously for research purposes. Website: https://bitsabout.me

Fairdrive Switzerland Awarded 2021 2022	Fairdrive is an open source decentralized personal data storage solution aiming to achieve user sovereignty over and privacy for their data. The peer-to-peer nature means no setting up of infrastructure is needed. There is no central authority to rely on or to block access to the data. The amount of storage used can be scaled as needed. This covers use cases for individuals storing their data as well as apps storing individual specific data, and others. By using Fairdrive integrated decentralized storage, developers can create and build interoperable, decentralized and open- sourced dApps so users can reclaim their privacy, own their data and control their digital identity. Faidrive stack is open sourced and developed under the Fair data society. Website: https://fairdrive.fairdatasociety.org/
Streamr Switzerland Awarded: 2020	Streamr (streamr.network) is the world's leading marketplace and decentralised network for real time data. The distributed, open-source, software project project was founded in 2017 with the mission of creating a platform to trade and distribute information, while allowing people to regain control of the data they produce. Through its Data Union concept, individuals can crowdsell their information on the Streamr network along with their fellow union members. Designed for safe data delivery and exchange, the Streamr network is scalable, low-latency and secure. Data Union members can monetize their google search history, what they've been shopping for on Amazon or even what type of coffee their smart coffee machine's been brewing. This data is extremely valuable for marketers, new market entrants, hedge funds, researchers and many more. Streamr believes that society will be better off when the value from, and control over, our data is decentralised away from a few giant corporations. Website: https://streamr.network
CANDIY South Korea Awarded: 2022	CANDIY is a new Web 3.0 based decentralized MyData operator that creates value for the individual user with their data in the following ways: 1. The CANDIY operator uses a blockchain to guarantee transparency and safety of data processing by documenting the user's consent and data provision history on the blockchain. 2. CANDIY introduces a method of disbursing blockchain tokens as rewards based on the participation level of each user. This tokenomics will eventually realize a web 3 world. 3. The source code of the CANDIY operator will be managed as an open-source repository within this year. We hope to attract developers across the MyData community to help build a transparent, secure, and interoperable ecosystem. If approved, the CANDIY operator will be the first to be awarded MyData Operator status in Korea. We aspire to decentralize our service and enter the EU market by the end of 2022. Website: https://candiy.io
Numbers Taiwan Awarded: 2020	Numbers provides data integrity assurance for data collection and exchange services. In order to better protect data privacy, many services today choose distributed data collection and storage such as keeping the information in personal mobile phones. This indeed enhances privacy and reduces the risks of transmitting data around. However, on the other hand, distributed data collection and storage require more trust between each party in order to maximise the value of data. Numbers helps service providers protect data integrity using blockchain and cryptographic technologies. At the time when data is generated, the environmental information such as time, location and other supporting metadata are captured and signed as a proof of the data. The hash value of the proof is registered on the blockchain to ensure that the user's data is kept intact and not modified. The integrity assurance provided by Numbers keeps the data private and distributed while still preserving its credibility. Website: https://numbersprotocol.io/

CitizenMe United Kingdom Awarded: 2021 2022	Data, Climate Change, Mental and Physical Health are the 4 greatest challenges facing humanity in the 21st Century and a post-pandemic world. Collectively, our digital Citizen Data must be used to help solve these global challenges. But, who owns our data? Currently 'Big Data' companies hoard, use and trade our data in unethical ways. We have no choice, no agency, no digital freedom. As digital citizens, we need a new way to interact with the internet. A way to collect and store our own data for ourselves. We need a way to interact with the always-on, online world - but without having to forever give away our data, for free. To change the way the internet works (for the better!) we've created CitizenMe. It's a breakthrough 'Zero Data' platform that focuses on the digital rights of us, the citizens of the internet. CitizenMe enables digital citizens to effortlessly collect their MeData and then choose how, where and when they exchange it, with liberty. Website: https://www.citizenme.com
DataYogi United Kingdom Awarded: 2020 2021 2022	DataYogi helps people unlock the power of their data. It does so by providing them with a secure data store and point of integration for their data, and the means to then co-manage this data with other parties (e.g. their suppliers). Website: https://www.datayogi.me
digi.me United Kingdom Awarded: 2020 2021	Digi.me facilitates individuals to share more & better data to enable businesses to provide more & better value, with 100% privacy, full security & consent. Individuals can get a full copy of their personal data (health, finance, social, wearables, media) every day, allows individual to store in personally encrypted location of their choice (OneDrive, Dropbox, Google Drive), and enables services to request access to extracts of that data through a Consent Certificate & SDK/API consent stack with a single call, and with full individual traceability with a Consent Dashboard. Importantly for individuals, only they hold their data, their data is fully secured and immutable. No data is shared without the individuals explicit consent and they have a consent dashboard to manage consents in a single place including revoking consent, audit logs, etc. The individual remains in full control at all times. Website: https://digi.me/
Pool United Kingdom Awarded: 2022	Pool is up-ending the current data economy by incubating and providing infrastructure for groups and developers building out applications and services that help ordinary people own and monetize their data. We call these data unions. The value proposition from a data union to its members is simple: Create, Share and Earn. But the technology needed to make this work at scale is complex. Pool provides a platform, governance and associated services to support data unions to develop, scale and monetize. At the top of this hierarchy sits the citizen. Through our Universal Data Wallet we provide personal storage and consent tools, sovereign identity management and payments – they stay in control. This drives the access data unions have to act in individuals best interests, when they represent their rights in a dispute or bargain collectively for their data's value. Organisations and scientists who need access to data insights can request products and query to an independent, raw data layer. Website: https://pooldata.io/
Mydex United Kingdom Awarded: 2022	The Mydex Charter provides a full description of our purpose and activities. Mydex Data Services Community Interest Company (Mydex CIC) is a social enterprise whose mission is to empower citizens with their data. Mydex provides citizens with personal data stores which enable them to collect, store, manage, use and share their own data under their own control independently of any organisation that may hold data about them. As a Community Interest Company under UK law, Mydex is legally required to always prioritise this mission. Everything about it - its technology, platform, business model, funding, business partners - is designed to align with and promote this mission. By making individuals the point at which data about themselves can be aggregated and integrated. Mydex provides them with a data asset that is theirs and that grows in richness and usefulness over their lives. Website: https://mydex.org

MyLife Digital United Kingdom Awarded: 2020	MyLife Digital has been built on the fundamental concept that individuals should be at the heart of how organisations collect, use and share their data. We follow both a data protection and privacy by design approach to ensure that our solutions are focussed on individuals first. By putting the individual first, organisations can build trust, whilst ensuring rigourous governance standards are maintained. Consentric by MyLife Digital, is a cloud-based platform that handles the governance of processing personal data. It delivers transparency to the individual on what data has been collected and for what purpose. It empowers the individual to make decisions over the use of their data across a number of data using services, and it delivers accountability and assurance via an immutable audit record of permissions granted or denied. Website: https://mylifedigital.co.uk/developer/
Self Innovations United States Awarded: 2020 2021 2022	At Self Innovations, we understand our well-being is about more than physical health. Our emotional and financial wellness, our accomplishments, even our relationship with the environment all contribute to Absolute Human Health. We focus on three services. 1. The development of the Self Framework. A blockchain and graph database hybrid that delivers the infrastructure for an application layer. 2. The Self Profile is a research-based categorical data structure and identity authentication tool based on self-sovereign identity and blockchain. 3. Software development for building on the Self Framework. Our current project, SelfPass, has launched with an open source and enterprise version. Users can log COVID-19 test results, vaccines, symptoms, quarantines, and treatment. Additionally, users can choose to anonymously share certain information with family and friends, and the medical community to help fight the pandemic. Website: https://selfinnovations.ai
Tru United States Awarded: 2020	Tru is a social publishing platform that will shortly emerge in public beta. It is primarily aimed at the USA market, and is a counter to 'fake news'. It is focused on the provenance of 'content' more so than the provenance of personal data, but as it is built with the JLINC protocol then personal data provenance is also a built in feature. Website: https://www.tru.net/

Appendix 2 – Proto-operators studied for the first edition of the paper

Smart Species Canada	Governance integration, WHiSSPR Auditors, Canadian OPN:Registrar, Smart Person, Smart City, Smart Nation. Consent DDE, Data Trust Governance for distributed transparency, DLC - digital ledger consent technology provider.
Peercraft Denmark	Currently a user-centric identity provider, Peercraft is working to become a purchasing agent for consumers via a fully decentralised business and service discovery protocol (opendiscovery.biz)
1001 Lakes Finland	1001 Lakes enables trusted data sharing for people and organisations to realise more value together.
City of Helsinki Finland	Helsinki wants to be the most functional city in the world by making full use of its data. Helsinki seeks to apply MyData principles in managing the personal data it collects and processes.
Findy Finland	The Findy consortium is working towards launching a collaboratively governed and operated public-private not-for-profit decentralised identity network.
Gravito Finland	Gravito is a cloud-based, real-time consumer profile which follows you automatically over domains and cross organisations. It allows you to define your domain specific multi-level consents and provides means to connect your profile to the surrounding device(s) and “things”. It gives organisations access to real-time consumer profiles/segments where the people are themselves communicating their preferences and consents.
Posti Finland	We at Posti believe in a fair, responsible, and transparent digital future. Embracing technologies and solutions that thrive the development towards a human-centric data economy should be the interest for every company as it is for us. Posti is the leading postal and logistics service company in Finland with over 22,000 employees. Posti manages the flow of everyday life by offering a broad range of postal, logistics, freight and e-commerce services.
Startup Commons Global Finland	Circle Pass is a service that is part of the ecosystemOS package provided by Startup Commons Global, focused on digitally connecting, visualising and benchmarking startup ecosystems for economic development and growth of entrepreneurship and innovation.
Younode Japan	Decentralised personal data sore which can work as a password manager also. Users can store it on their own device or Google Drive that you can manage.
Holland Health Data Co-operative Netherlands	HHDC empowers its members (citizens) with an ethical check on requests for use of individual health data, based on the consent structure they have specified.
Qiy Foundation Netherlands	Co-creation with market parties of a trust-based human-centric online ecosystem with individuals as a constitutional part in control over their data.
Healthbank cooperative Switzerland	The global people-owned platform for managing your health and medical data in one secure database.
MIDATA Switzerland	MIDATA Cooperative has established a governance model and IT platform solution for citizen-centred and patient-centred health data aggregation, allowing citizens and patients to give dynamic and granular consent to data use. The MIDATA platform embodies modern data governance principles, enabling health research and health services, while at the same time ensuring citizens’ and patients’ sovereignty over their personal data. The platform is based on advanced database and encryption technologies developed at ETH Zurich. Its FHIR API enables interoperability and use of structured data. The platform acts as a hub for a mobile app ecosystem. The platform and app framework are operational and being further developed in the context of the SPHN initiative and further national initiatives.

Consentua United Kingdom	Consentua lets organisations orchestrate their data processing based on the consent that they have from data subjects. Consentua collects, stores and updates consent records so that business processes can be automatically started and stopped, and provides a rich audit trail of consent collection and use.
Dataswift United Kingdom	Dataswift is a technology company that develops data portability and processing tools leveraging the Hub of All Things (HAT) personal data account, enabling individuals and businesses to implement and benefit from the ethical storage and processing of data.
Hub-of-All-Things United Kingdom	The HAT Community Foundation is devoted to advancing the Hub-of-All-Things (HAT) open source technology, and to advancing the interests of HAT owners everywhere. It acts as regulator for the HAT ecosystem, and operates the HAT-LAB, which functions as the research and innovation centre for the HAT.
Powr of You United Kingdom	Powr of You is a consumer data hub helping people make money from their data, with actual behaviour data from mobile, browsers, social, lifestyle apps.
HIE of One United States	HIE of One Trustee is a standards-based, Free / Open Source software suite for substitutable operators with decentralised governance. We use health information exchange (HIE) as the use case.
Indie Computing United States	Your data on hardware you control. Indie Computing provides managed appliances to enable consumers, families, and organisations to manage their valuable data in place they control.
JLINC United States	JLINC is a protocol for permissioned data sharing that enables multiple parties to co-manage data assets in a human-centric way.
Prifina United States	Prifina is a user-held data platform that provides tools for developers to build direct-to-consumer applications and widgets, on superior data that never leaves the individual.
Spartacus United States	Spartacus was incorporated in 2019 as Data Fiduciary Inc. We help our customers take back their privacy and protect their data and identity.
UBDI United States	UBDI allows individuals to securely aggregate millions of data points about themselves from their social, financial, wearable, and health accounts and get paid for their time and attention when seeing relevant ads or for sharing insights from that data for market and financial research.

Appendix 3 – Technologies, specifications and standards commonly in use

This list of technologies, specifications and standards is maintained as an online resource at: https://mydata-global.org/operator-standards

Decentralized Identifiers (DIDs)	Identity management
Decentralized identifiers (DIDs) are a type of identifier that enables a verifiable, decentralized digital identity. A DID refers to any subject (e.g., a person, organisation, thing, data model, abstract entity, etc.) as determined by the controller of the DID.	
DIDComm Messaging	Identity management
DIDComm is a set of tools to allow horizontal and bidirectional channels of communication between two entities that know each other's DIDs and nothing else.	
eIDAS	Identity management
eIDAS (electronic IDentification, Authentication and trust Services) is an EU regulation on electronic identification and trust services for electronic transactions in the European Single Market. All organisations delivering public digital services in an EU member state must recognize this electronic identification from all EU member states.	
Identity Mixer (Idemix)	Identity management
IBM Identity Mixer is a cryptographic protocol suite for privacy-preserving authentication and transfer of certified attributes. It allows user authentication without divulging any personal data.	
Hyperledger Indy	Identity management
Hyperledger Indy provides tools, libraries, and reusable components for providing digital identities rooted on blockchains or other distributed ledgers so that they are interoperable across administrative domains, applications, and any other silo.	
National eID	Identity management
An electronic identification is a digital solution for proof of identity of citizens or organisations. They can be used to view and access benefits or services provided by government authorities, banks or other companies, for mobile payments, etc. Apart from online authentication and login, many electronic identity services also give users the option to sign electronic documents with a digital signature.	
OAuth 2.0	Identity management
OAuth (Open Authorization) is an open standard for access delegation, commonly used as a way for Internet users to grant websites or applications access to their information on other websites but without giving them the passwords. This mechanism is used by companies to permit the users to share information about their accounts with third-party applications or websites.	
OpenID Connect (OIDC)	Identity management
OpenID Connect is an authentication layer on top of the OAuth 2.0 authorization framework. It allows computing clients to verify the identity of an end-user based on the authentication performed by an authorization server, as well as to obtain the basic profile information about the end user in an interoperable and REST-like manner.	
SAML 2.0	Identity management
Security Assertion Markup Language (SAML) is an open standard for exchanging authentication and authorization data between parties, in particular, between an identity provider and a service provider. SAML is an XML-based markup language for security assertions (statements that service providers use to make access-control decisions).	

Self-Issued OpenID Provider (SIOP)	Identity management
This specification extends OpenID Connect with the concept of a Self-Issued OpenID Provider (Self-Issued OP), an OP which is within the End-User's local control. End-Users can leverage Self-Issued OPs to authenticate themselves and present claims directly to the RPs. This allows users to interact with RPs directly, without relying on third-party providers or requiring the End-User to operate their own hosted OP infrastructure.	
System for Cross-domain Identity Management (SCIM)	Identity management
System for Cross-domain Identity Management (SCIM) is a standard for automating the exchange of user identity information between identity domains, or IT systems.	
Principles of Self Sovereign Identity (SSI)	Identity management
Twelve principles represent the SSI community's consensus on the core ethics and policies that must guide any digital identity ecosystem that chooses to align with SSI.	
FIDO Universal 2nd Factor (U2F)	Identity management
Universal 2nd Factor (U2F) is an open standard that strengthens and simplifies two-factor authentication (2FA) using specialised Universal Serial Bus (USB) or near-field communication (NFC) devices based on similar security technology found in smart cards.	
FIDO Web Authentication (WebAuthn)	Identity management
Web Authentication (WebAuthn) is a core component of the FIDO2 Project with the goal to standardise an interface for authenticating users to web-based applications and services using public-key cryptography.	
Universal DID Resolver	Identity management
The Universal Resolver resolves Decentralized Identifiers (DIDs) across many different DID methods, based on the W3C DID Core 1.0 and DID Resolution specifications. It is a work item of the DIF Identifiers & Discovery Working Group.	
User-Managed Access (UMA)	Identity management
User-Managed Access (UMA) is an OAuth-based access management protocol standard.	
Verifiable Credentials (VC)	Identity management
Verifiable Credentials provide a mechanism to express credentials on the web in a way that is cryptographically secure, privacy respecting, and machine-verifiable.	
Consent receipts	Permission management
Consent receipts are a way to record a given consent in a standardised way. Having a consent receipt is good for both the individual as well as for the data controller. It is a record of agreement about usage of PII that both parties can refer to.	
CRUD / CRUDS	Permission management
In computer programming, create, read, update, and delete (CRUD) are the four basic operations of persistent storage. Share operation can be considered as the fifth basic operation.	
Data Privacy Vocabulary (DPV)	Permission management
The Data Privacy Vocabulary (DPV) provides terms (classes and properties) to describe and represent information related to processing of personal data based on established requirements such as for the EU General Data Protection Regulation (GDPR).	
Hyperledger Aries RFC 0167: Data Consent Lifecycle	Permission management
A reference implementation for generating a consent proof for use with DLT (Distributed Ledger Technology).	
IEEE P7012 - Machine Readable Personal Privacy Terms	Permission management
The standard identifies/addresses the manner in which personal privacy terms are proffered and how they can be read and agreed to by machines.	

ISO 27560 Consent record information structure	Permission management
The standard will include guidelines for using consent receipts and consent records associated with a PII Principal's data processing consent to help support: Providing the PII Principal with a record of the consent; exchanging consent information between various information systems; and maintenance of the recorded consent throughout its lifecycle. In the standard, neither receipts nor records will be exchanged, nor will the exact structure of such exchanges be specified.	
ISO 29184 Online privacy notices and consent	Permission management
This document specifies controls which shape the content and the structure of online privacy notices as well as the process of asking for consent to collect and process personally identifiable information (PII) from PII principals.	
JLINC protocol	Permission management
JLINC is an open protocol for sharing data protected by an agreement on the terms under which the data is being shared.	
Kantara consent receipt	Permission management
A Consent Receipt is a record of consent used by a PII Controller as their authority to collect, use and disclose a PII Principal's personally identifiable information (PII).	
The Open Digital Rights Language (ODRL)	Permission management
The Open Digital Rights Language (ODRL) is a policy expression language that provides a flexible and interoperable information model, vocabulary, and encoding mechanisms for representing statements about the usage of content and services.	
Web Distributed Authoring and Versioning (WebDAV)	Permission management
WebDAV (Web Distributed Authoring and Versioning) is an extension of the Hypertext Transfer Protocol (HTTP) that allows clients to perform remote Web content authoring operations.	
Common Core Ontologies (CCO)	Data model management
INCITS 573-2 Common Core Ontology is a set of terms, definitions, and relations commonly used across multiple domains, which will enable conforming extensions for specific domains or applications. CCO conforms to Basic Formal Ontology (BFO) and includes a subset User Profile Ontology.	
Fast Healthcare Interoperability Resources (FHIR)	Data model management
Fast Healthcare Interoperability Resources (FHIR, pronounced "fire") is a standard describing data formats and elements (known as "resources") and an application programming interface (API) for exchanging electronic health records (EHR).	
ISO 7250 (human body measurements)	Data model management
The standard provides a description of anthropometric measurements which can be used as a basis for comparison of population groups and for the creation of anthropometric databases.	
ISO 21838 Basic Formal Ontology (BFO)	Data model management
Basic Formal Ontology (BFO) is a small, upper level ontology that is designed for use in supporting information retrieval, analysis and integration in scientific and other domains.	
JSON-LD	Data model management
JSON-LD is a lightweight syntax to serialise Linked Data in JSON [RFC8259]. Its design allows existing JSON to be interpreted as Linked Data with minimal changes. JSON-LD is primarily intended to be a way to use Linked Data in Web-based programming environments, to build interoperable Web services, and to store Linked Data in JSON-based storage engines.	
Linked data	Data model management
Linked data is structured data which is interlinked with other data so it becomes more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers.	
Open Data Protocol (OData)	Data model management
Open Data Protocol (OData) is an open protocol that allows the creation and consumption of queryable and interoperable REST APIs in a simple and standard way.	

Open Database Connectivity (ODBC)	Data model management
Open Database Connectivity (ODBC) is a standard application programming interface (API) for accessing database management systems. An application written using ODBC can be ported to other platforms, both on the client and server side, with few changes to the data access code.	
Hedera Hashgraph	Personal data transfer
Hyperledger Aries is infrastructure for blockchain-rooted, peer-to-peer interactions. It defines messaging protocols and implements those protocols in shared, reusable, interoperable tool kits designed for initiatives and solutions focused on creating, transmitting and storing verifiable digital credentials.	
Hyperledger Aries	Personal data transfer
Hyperledger Aries is infrastructure for blockchain-rooted, peer-to-peer interactions. It defines messaging protocols and implements those protocols in shared, reusable, interoperable tool kits designed for initiatives and solutions focused on creating, transmitting and storing verifiable digital credentials.	
Lightweight Directory Access Protocol (LDAP)	Personal data transfer
The Lightweight Directory Access Protocol (LDAP) is an open, vendor-neutral, industry standard application protocol for accessing and maintaining distributed directory information services over an Internet Protocol (IP) network.	
IOTA Masked Authenticated Messaging (MAM)	Personal data transfer
Masked Authenticated Messaging (MAM) is a second layer data communication protocol which adds functionality to emit and access an encrypted data stream, like RSS, over the Tangle (IOTA's distributed ledger) regardless of the size or cost of device. IOTA's consensus protocol adds integrity to these message streams.	
Transport Layer Security (TLS)	Personal data transfer
A cryptographic protocol designed to provide communications security over a computer network, successor of SSL (Secure Sockets Layer). Use of the TLS protocol as the security layer in HTTPS remains the most publicly visible. Latest, currently recommended version of the TLS protocol is TLSv1.3.	
The WebSocket Protocol	Personal data transfer
WebSocket is a computer communications protocol, providing full-duplex communication channels over a single TCP connection. The WebSocket protocol was standardised by the IETF as RFC 6455 in 2011, and the WebSocket API in Web IDL is being standardised by the W3C.	
Confidential Storage	Personal data transfer
A privacy-respecting mechanism for storing, indexing, and retrieving encrypted data at a storage provider.	
Role-Based Access Control (RBAC)	Personal data transfer
In computer systems security, role-based access control (RBAC) or role-based security is an approach to restricting system access to authorised users. RBAC is a policy-neutral access-control mechanism defined around roles and privileges.	
Immutable logging	Logging and accountability
An immutable audit log is a tamper-resistant recording of how a system has been used.	
Advanced Encryption Standard (AES)	General
The Advanced Encryption Standard (AES) is a specification for the encryption of electronic data established by the U.S. National Institute of Standards and Technology (NIST) in 2001.	
Distributed Ledger Technology (DLT)	General
A distributed ledger (also called a shared ledger or distributed ledger technology or DLT) is a consensus of replicated, shared, and synchronized digital data geographically spread across multiple sites, countries, or institutions. Unlike with a centralized database, there is no central administrator.	

ISO 9001	General
The ISO 9000 family of quality management systems (QMS) is a set of standards that helps organizations ensure they meet customer and other stakeholder needs within statutory and regulatory requirements related to a product or service.	
ISO 27001	General
ISO/IEC 27001 is an international standard on how to manage information security. It sets out the specification for an information security management system (ISMS) - a best-practice approach helps organisations manage their information security by addressing people, processes and technology.	
ISO 27007	General
The standard provides guidance on managing an information security management system (ISMS) audit programme, on conducting audits, and on the competence of ISMS auditors. This standard is applicable to those needing to understand or conduct internal or external audits of an ISMS or to manage an ISMS audit programme.	
ISO 27017	General
A security standard developed for cloud service providers and users to make a safer cloud-based environment and reduce the risk of security problems.	
ISO 27018	General
A security standard to help cloud service providers who process personally identifiable information (PII) to assess risk and implement controls for protecting PII.	
JSON Web Token (JWT)	General
JSON Web Tokens (JWT) are an open, industry standard RFC 7519 method for representing claims securely between two parties. JWT.IO allows you to decode, verify and generate JWT.	
RSA Cryptography Specifications	General
RSA (Rivest–Shamir–Adleman) is a public-key cryptosystem that is widely used for secure data transmission.	
Secure Hash Algorithms (SHA)	General
The Secure Hash Algorithms (SHA) are a family of cryptographic hash functions published by the National Institute of Standards and Technology (NIST) as a U.S. Federal Information Processing Standard (FIPS).	
Swagger	General
Swagger is an Interface Description Language for describing RESTful APIs expressed using JSON. Swagger is used together with a set of open-source software tools to design, build, document, and use RESTful web services.	
Trusted Execution Environment (TEE)	General
A trusted execution environment (TEE) is a secure area of a main processor (for example Arm Trust-zone). In general terms, the TEE offers an execution space that provides a higher level of security for trusted applications running on the device than a rich operating system.	
X.509 Public Key Infrastructure Time-Stamp Protocol (TSP)	General
The Time-Stamp Protocol, or TSP is a cryptographic protocol for certifying timestamps using X.509 certificates and public key infrastructure. The timestamp is the signer's assertion that a piece of electronic data existed at or before a particular time.	
X.509 Public Key Infrastructure Certificate	General
In cryptography, X.509 is a standard defining the format of public-key certificates. X.509 certificates are used in many Internet protocols, including TLS/SSL, which is the basis for HTTPS, the secure protocol for browsing the web. They are also used in offline applications, like electronic signatures.	

Citation: Langford, J., Poikola, A., Janssen, W., Lähteenoja, V. and Rikken, M. (Eds.) (2022) '*Understanding MyData Operators*', MyData Global.

Graphic design: Kirmo Kivelä