Privacy-friendly data storage for collectives

A study to investigate the use of Personal data vaults for citizen collectives performed by TNO in collaboration with Stichting Je Leefstijl Als Medicijn

TNO Public ⟩ TNO 2025 R12932

12 december 2025

ICT, Strategy & Policy
www.tno.nl

TNO 2025 R12932 – 12 december 2025

Privacy-friendly data storage for collectives

A study to investigate the use of Personal data vaults for citizen collectives performed by TNO in collaboration with Stichting Je Leefstijl Als Medicijn

Authors	Willem Datema, Cornelis Bouter, Kit Buurman

Classification report	TNO Public
Title	TNO Public
Report text	TNO Public
Amount of pages	16 (excl. front and back cover)
Amount of appendices	0
Project name	ERP Better Together
Project number	060.64033

Table of Contents

1 Introduction

Citizen collectives are increasingly seen as essential for needed innovation and transitions, to tackle great societal challenges in sustainability and in the health domain. These challenges ask for new ways of working together, research and innovation, where co-creation is vital. Within the Better Together project, TNO develops new methods to facilitate collaboration between citizens, the government, market and knowledge institutes. TNO brings independent knowledge that enables systemic changes, while collaborating closely with several citizen collectives in living labs in the energy, health and social cohesion domains.

One of these citizen collectives is called 'Stichting Je Leefstijl als Medicijn (JLAM)'. This is a collective that was founded to motivate people to live healthier by making changes to their lifestyle. They inform about lifestyle and health, organize platforms like support groups where experts by experience help and motivate each other. They are successful in this, since they report 700.000 page visits and 21.000 people in peer support groups. The organization consists of a central team and 59 volunteers and they are very active in partnerships as well. One of the groups that is part of the JLAM foundation is called Diabetes2Doorbreken. The group has been part of a pilot study done by TNO Healthy Living in 2020[1], which researched community factors and feasibility of gathering and sharing health data. The outcome was positive: 80 percent of the respondents said they achieved more with the community than they could have done alone.

Part of the study was the possibility to share measurements with each other every week. With these measurements, coaches help participants to guide them towards a better lifestyle, based on their own experiences[2]. This is called 'Zaterdag Wegen en Meten (ZWEM)' which translates to 'Weighing and Measuring on Saturday', and this initiative continues to this day. 3100 participants were part of this initiative in 2025, which is why the need for better data entry increased. Therefore, the ZWEM application was created to be able to enter 5 measurements on each Saturday. This system works well, but like many other applications, it requires people to enter their sensitive medical data to a central database, after which they have no control over it anymore. Therefore, TNO and JLAM together investigated a way of combining the ease of use of the ZWEM application in a more privacy-friendly manner.

A technology that can take privacy aspects into account is the technology of personal data vaults. A personal data vault is a place where data of various applications is stored in one place, granting full ownership of the data to the holder of the personal data vault. Applications become clients to the data, with the owner being in control on whether they would like to grant access to the data or not. Various architectures around personal data vaults exist, but the Solid[3] protocol is the most known. This protocol focuses on interoperability, is based on web standards, and puts people in control over their own data. The personal data vaults in the Solid world are known as Pods. An added benefit of the Pods is that data that is stored in such a pod can be reused between applications. This means the user only has to enter their weight once, and multiple applications have access to this data point. For this to work,

[1] https://www.jeleefstijlalsmedicijn.nl/tno-onderzoek-bevestigt-je-leefstijl-als-medicijn-werkt/
[2] https://www.jeleefstijlalsmedicijn.nl/op-weg-naar-optimale-leefstijlcoaching-hoe-tno-ervaringsdeskundigen-inzet/
[3] https://solidproject.org/

semantic interoperability is required so that each application knows what each data point means.
In this work, we explore the possibility of moving from the centralized ZWEM application to a decentralized network of personal data vaults, while also focusing on data interoperability to facilitate data reuse.

1.1 Research question

The aim of this work was to enable data reuse by focusing on privacy and semantic interoperability. This led to the following research question: “How can personal data vaults, specifically Solid pods, enable privacy-preserving data storage within citizen collectives while enabling data reuse across applications?”
This research explores whether centralized databases can be replaced with decentralized personal data storage in a real-world health application, without sacrificing the simplicity and user experience that people expect from this web application. We investigate:

1. Technical feasibility: “Can health data be stored in personal data vaults while keeping the application easy to use?”
2. User control and Privacy by design: “Can users maintain full ownership and control over their sensitive health information?"
3. User experience: “Can the current state of the Solid provide data insertion methods that do not overburden the user?”
4. Semantic Interoperability: “Can we used standardized health vocabularies to enable data sharing between different applications?”

This work was done together with the volunteers of the JLAM foundation. We did this by using the actual ZWEM application developed by the JLAM foundation, however we did *not* use any personal data of the participants.

2 Implementation

This section describes the implementation part of the study. It starts with the user stories that were created at the start of the study and is followed by the architecture description. Afterwards, the proof of concept and the standards that were used are explained. Finally, the discussion and future directions are given.

2.1 Use cases / User stories

To decide what functionalities should be changed to the application as-is, we came up with the following user stories:

2.1.1 User Stories #1: From the User perspective

- “As a participant in ZWEM, I want to track the progress of my health gains and—if I give my consent—share this data with the ZWEM organizer in a way that protects my privacy.”
- I take my measurements at home and store them in my personal data vault (whether it already exists or not).
- I can view the history of my measurements, including a visualization of how the measurements have developed over time.
- I can—if and as long as I give my consent—share this data with the ZWEM organizer in a way that protects my privacy. I can give separate consent for use of (a selection of) my individual measurement data, consisting of weight, diabetes blood values, waist circumference, height, age and gender.
- I can stop participating in ZWEM at any time, and then my data will no longer be shared. Aggregated data may still be retained by the ZWEM organizer for (effectiveness) research.

2.1.2 User story #2: From the perspective of the collective (JLAM)

- “As a ZWEM organizer, I want to be able to aggregate users’ data in order to demonstrate the effectiveness/health benefits of the ZWEM initiative as a whole (e.g., for scientific research).”

2.2 Architecture

This section explains how the architecture of the ZWEM application has transformed to offer control, privacy and ownership of their health information to users, while maintaining the simplicity that users expect.

2.2.1 Initial state of the application

The initial state of the ZWEM application was like most web applications people use daily. All user data was stored in a central database controlled by the application.

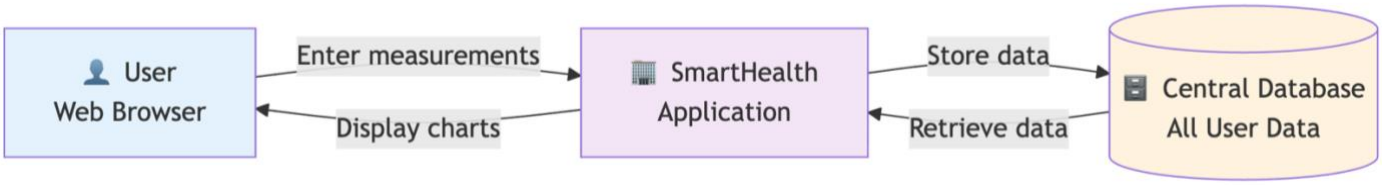

Figure 1: Architecture of the initial state of the application

Figure 1 displays the flow of the application. The user logs in to the ZWEM application and enters their measurements. This data is then stored in the central database where it can be retrieved by the application to display charts.
Limitations of this architecture are that users do not have direct control over their data, since their health information is stored on the servers of the ZWEM application. Furthermore, their data is difficult to be reused by other health applications, even though the measurements are suited for reuse very well. Additionally, a user needs to trust the owners of the ZWEM application, potentially blocking them from entering their sensitive information in case the owners are unknown or not trusted by the user. Lastly, removal of the data can only happen at request of the user, they have no way to verify whether their data is actually removed.

2.2.2 The new approach

Given the limitations highlighted in the previous section, a new architecture was developed. This architecture uses a hybrid model, where users can choose whether their data is saved in the central database or in their personal data vault. We integrated support for Solid personal online data stores (pods) as implementation of this personal data vault concept. Figure 2 displays the two options and the place where the data is stored. This Solid pod can be hosted at different places. The most obvious one is at a pod provider, a third party that offers data vaults as a service. The open protocol of Solid allows for an easy switch between pod providers and creates a level playing field, so it becomes easier for potential new parties to enter the market. One could even host their own Solid pod, ensuring no other parties have access to their private data.

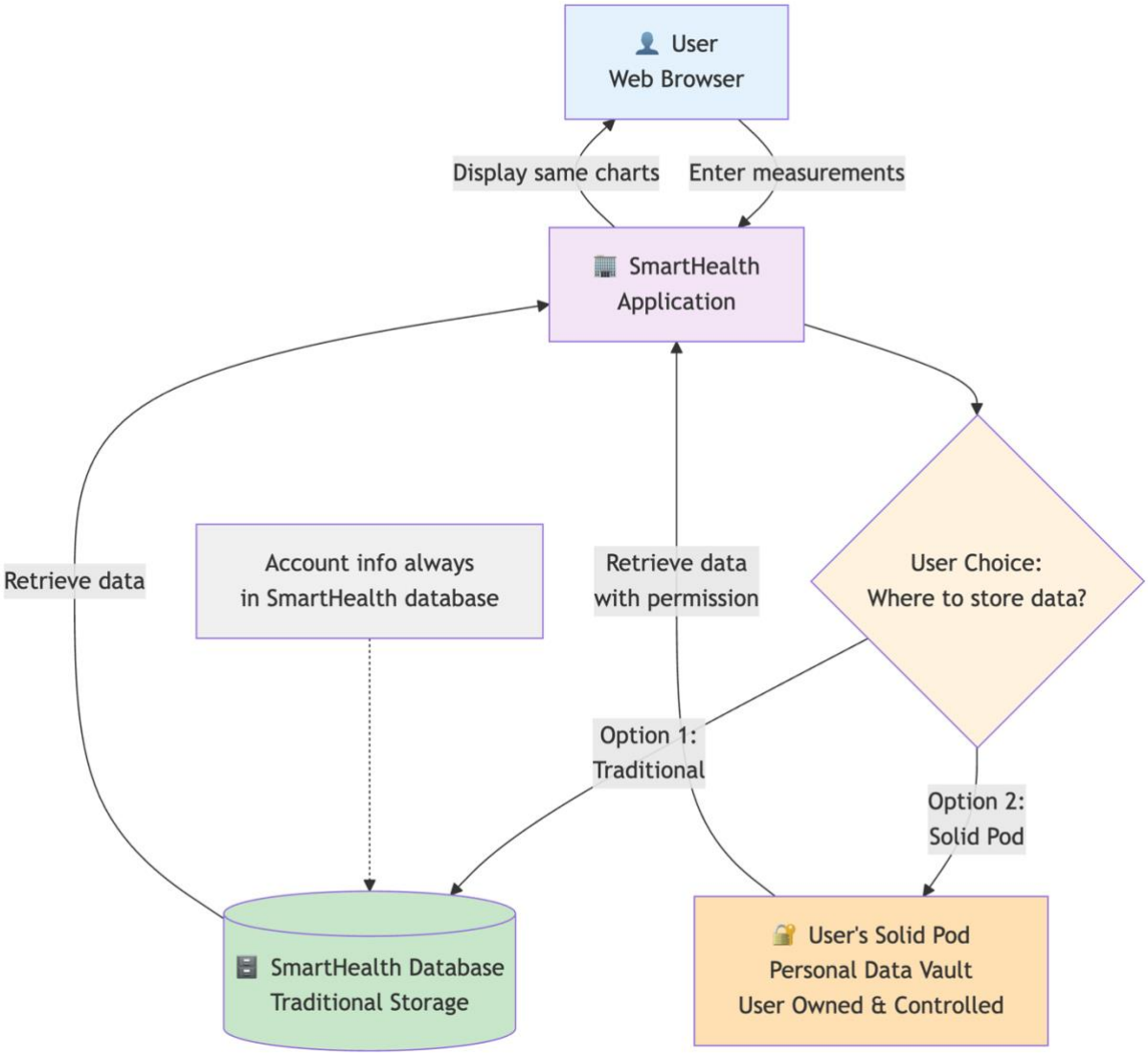

Figure 2: Current architecture of the ZWEM application

The application now knows two modes, depending on what each user chooses. Option 1 is exactly the same as the initial state where the data is stored in the central database. Option 2 is the new part of the architecture, where data is saved in the personal data vault of a user. As explained before in the User Stories, it was important that the user experience would not change even though the data was saved in another place. The only difference with this architecture is where the data lives and who controls it, interface towards the user remains almost identical.

2.3 Proof-of-Concept

The architecture explained above was implemented in a proof-of-concept implementation. The ZWEM application was taken as a starting point, and features were added over the span of a couple of months. For the operations on Solid pods, the client libraries[4] provided by Inrupt[5] were used. These libraries offer standardized operations based on the Solid protocol,

[4] https://docs.inrupt.com/sdk/javascript-sdk
[5] https://www.inrupt.com/about

and are created by Inrupt, the company that the founder of Solid (Sir Tim Berners-Lee) founded. Figure 3 displays all the components used in the proof-of-concept implementation.

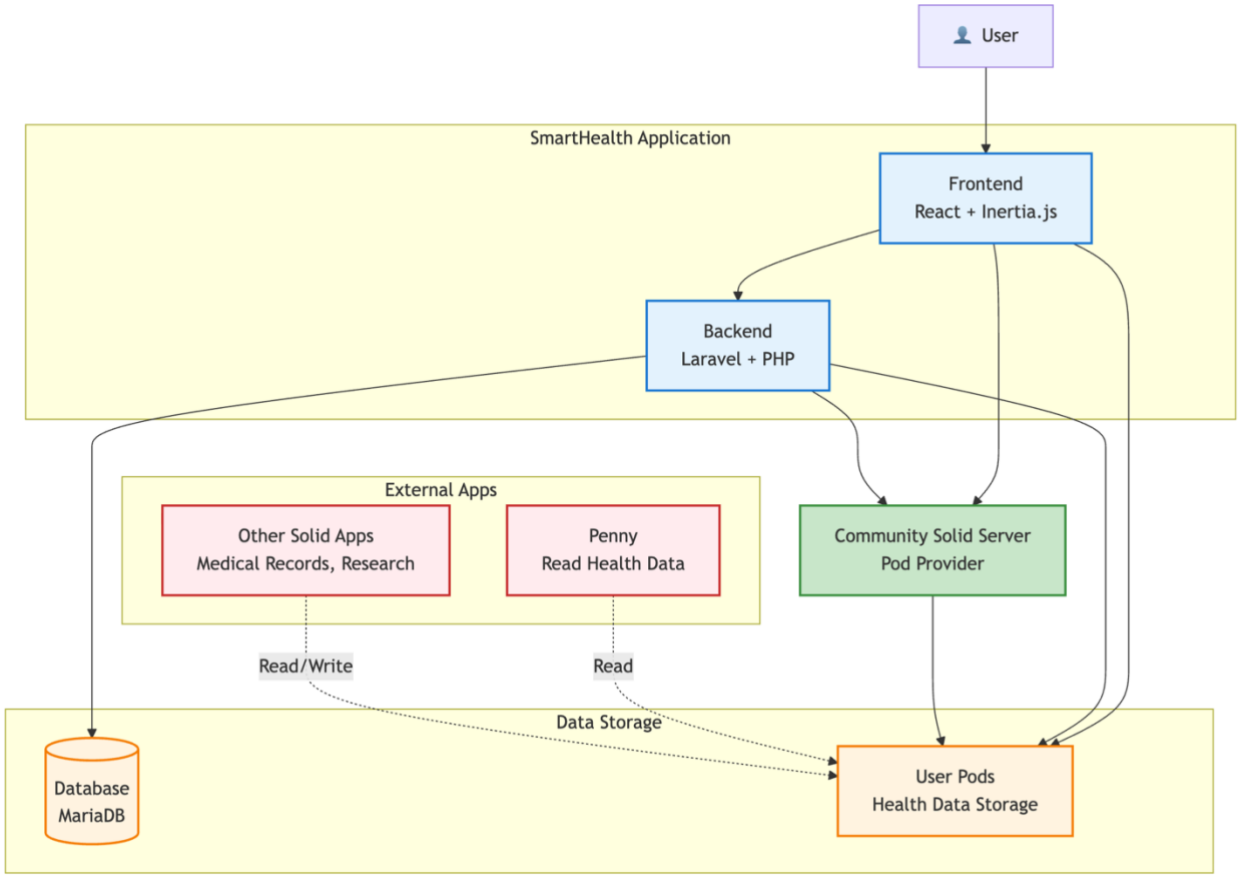

Figure 3: Component diagram of the proof-of-concept implementation

The user interacts with the frontend and can manage their Solid access from there; no requests go via the backend to ensure data only flows between the solid pod and the browser. They login to the ZWEM application first, and login to the Community Solid Server[6] afterwards. The Community Solid Server is an open-source implementation of the Solid protocol, which was used for this study. The second login is needed to manage access and to save measurements. After the measurements are saved, they are open to be read by other applications that implement the Solid protocol, or to general data viewing applications like Penny[7].

As there is a requirement that says that there should not be many changes in user experience, there is an option to create a Solid pod in the ZWEM application, in case the user does not have a Solid pod yet. This can be done on registration in the ZWEM application, or at any time via the Profile section. The app creates the pod and the necessary files for the ZWEM application to operate well, and the user just needs to login and give the ZWEM application access to the pod.
Next to this, we also showed that existing pods can be used by the ZWEM application. We created another application (SmartScale) that saves weight, as a demo for a smart scale.

[6] https://communitysolidserver.github.io/CommunitySolidServer/latest/
[7] https://penny.vincenttunru.com/

When a user returns to the ZWEM application and logs in with their Solid pod, the data from the other application will show up, demonstrating data interoperability.
Another feature that was added to the ZWEM application was the opt-out of general statistics. The ZWEM application has a screen that shows how much weight was lost by all participants. This number is computed over multiple users in the ZWEM database (for users that do not use a personal data vault) and all the Solid pods combined. In case a user does not want their data to be used, they can easily revoke access in their Profile section. The data will not be able to be used any more, since the access control changes in such a way that the ZWEM application cannot access it anymore. Figure 4 shows the Profile section with switches to grant or revoke access. Note that this can be done because there is a difference between the user accessing data and the application accessing data. A user can always see their data in the ZWEM application because data are only retrieved by the frontend, while the general statistics feature requires the backend of the ZWEM application to access the data, which can be allowed or prohibited by the user.

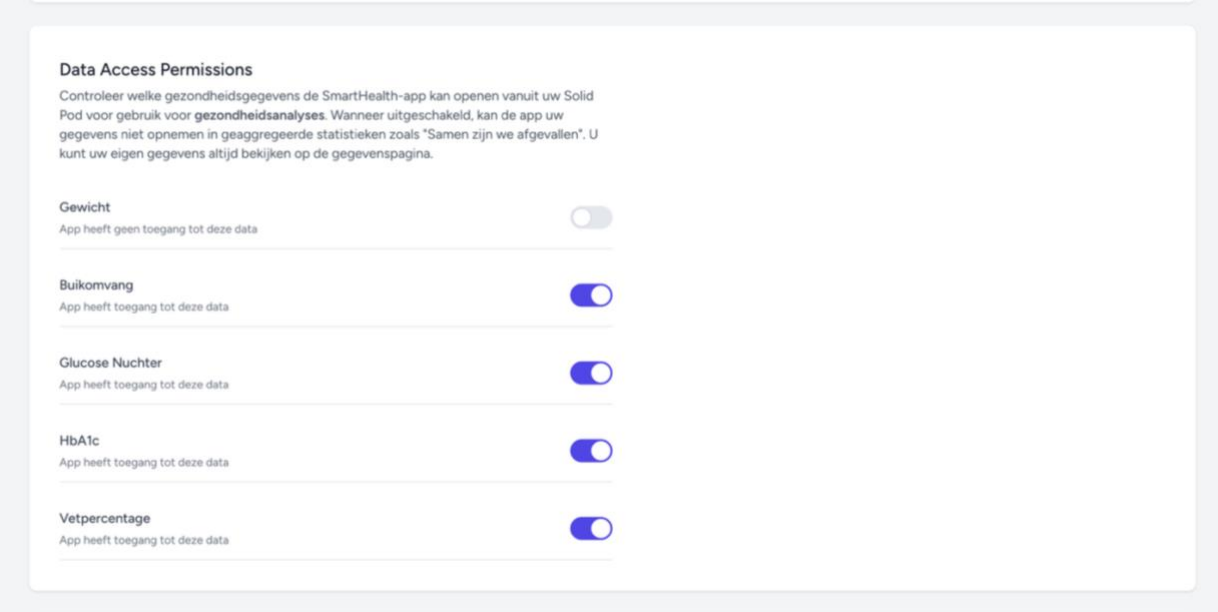

Figure 4: A screenshot of the user permissions for general statistics. The user has given access to every measurement except his weight.

2.4 Standards

To ensure data that are generated in the ZWEM application can be used by other applications, we adopted the reusability principles from the field of semantic interoperability. The Solid technology is rooted in the same Linked Data technology. In practice, this means we reuse existing standardized models from the healthcare domain and from the domain of observations and measurements, because the combination of those two domains covers the data inserted into the ZWEM application.
The healthcare standards that were used for the PoC are Fast Healthcare Interoperability Resources (FHIR[8]) and Logical Observation Identifiers Names and Codes (LOINC[9]). FHIR is an international standard developed by healthcare IT experts for exchanging electronic health

[8] https://www.fhir.org/
[9] https://loinc.org/

records. It contains the basic structure for storing health observations. This template contains, for example, the date, a value, who it belongs to and what was measured. This is combined with LOINC, which tells us something about semantics. LOINC is a universal catalog of medical and laboratory observations with unique codes for each type of measurements. The codes that are used are as follows:

LOINC Code	**Name**
29463-7	Body Weight
56086-2	Waist Circumference
2339-0	Blood Glucose
4548-4	Hemoglobin A1c (HbA1c)
41982-0	Body Fat Percentage

These codes are used by many other organizations and ensure that we talk about the same properties. Instead of calling it ‘weight’ or ‘body mass’, we use a specific code that means exactly the same everywhere in the world. Another aspect that requires standardization is the units of measurement. For this we use the Quantities, Units, Dimensions and Data Types ontology (QUDT). This is a standard vocabulary for units of measurements. This prevents confusion between, for example, pounds and kilograms. Another vocabulary that can be used to express the measurements is the SNOMED Clinical Terminology (CT) vocabulary. This provides codes for clinical context. For example, code 264362003 means “home environment” which tells us where the measurement was taken.
We additionally use several domain-independent standardized models that model the observations and measurements ubiquitous in a wearable or Internet of Things environment. The Open Geospatial Consortium (OGC) has produced the widely respected Observations, Measurements and Samples standard (O&M). The Semantic Sensor Network ontology (SSN/SOSA) and the SAREF ontology (Smart Applications Reference Ontology) have been developed as semantic implementations of O&M. The SAREF ontology has since been adopted by the European Commission for standardization of home appliances. Lastly, the Friend of a Friend ontology (FOAF) is used to represent the person the measurements concern.
The conceptual model developed in these standards divides a measurement into several aspects (see also Figure 5):

1) The sensor that recorded the measurement, e.g. a scale used to weigh yourself.
2) The (observable) property that was recorded, e.g., your weight.
3) The thing in the real world we measured the property of, e.g., “Willem Datema”.
4) The value that was recorded, e.g., 73.5.
5) The unit of measure that the value is in, e.g., kilogram.
6) The moment the measurement was made, e.g, noon on September 30th 2025.

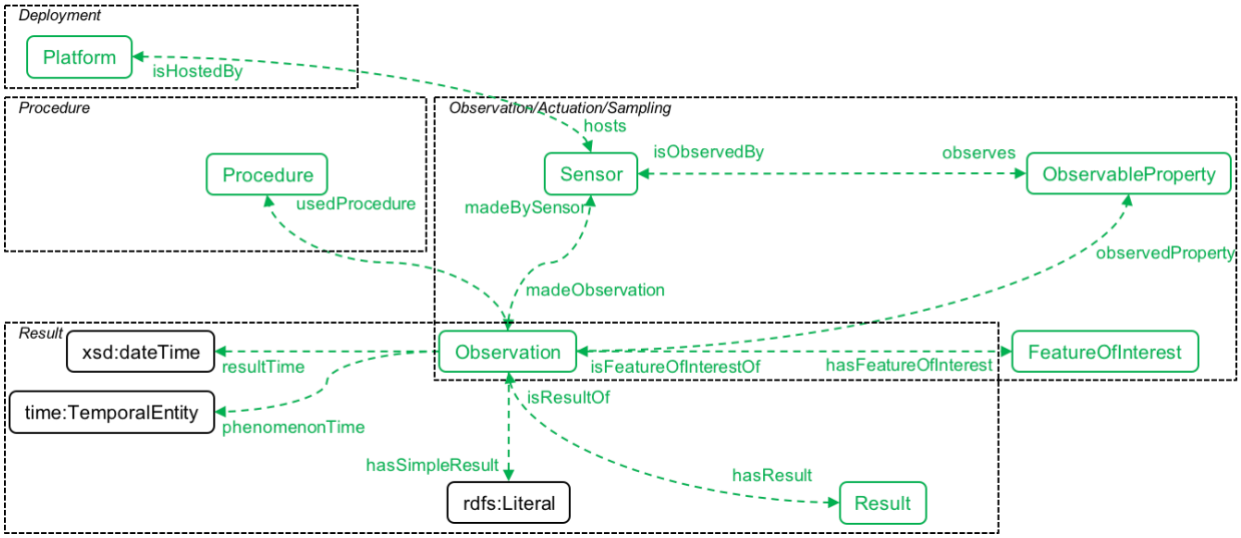

Figure 5 Overview of main SOSA/SSN classes (source: W3C SSN vocab)

In the European IHI project iCare4CVD[10] we have worked on aligning the healthcare models with the wearable models. This resulted in a work-in-progress Digital Biomarker Ontology (DBO) that reuses the foundation of SOSA/SSN with an alignment to the healthcare vocabularies (Figure 6). In Figure 7 an example body weight measurement is modelled according to the DBO, showing a clear distinction between the classes used for generic wearable readings, used for the wearer involved, and used to refer to a medical taxonomy. In the proof of concept, we have implemented a translation from the tabular source data to this shared ontology, such that all measurements entered via the existing interface are translated to the enriched common format without an additional burden on the user. The Solid technology is founded on the same linked data principles, so the Solid libraries provided by Inrupt offer native reusable classes for this purpose.

[10] https://icare4cvd.eu

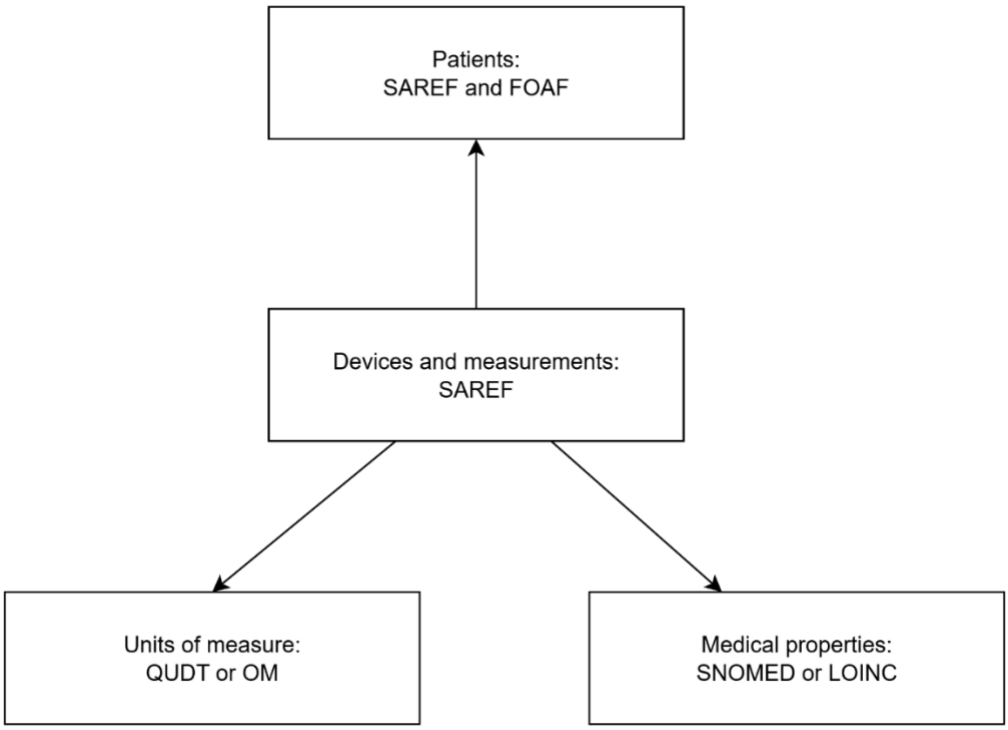

Figure 6 Digital Biomarker Ontology architecture or modularisation

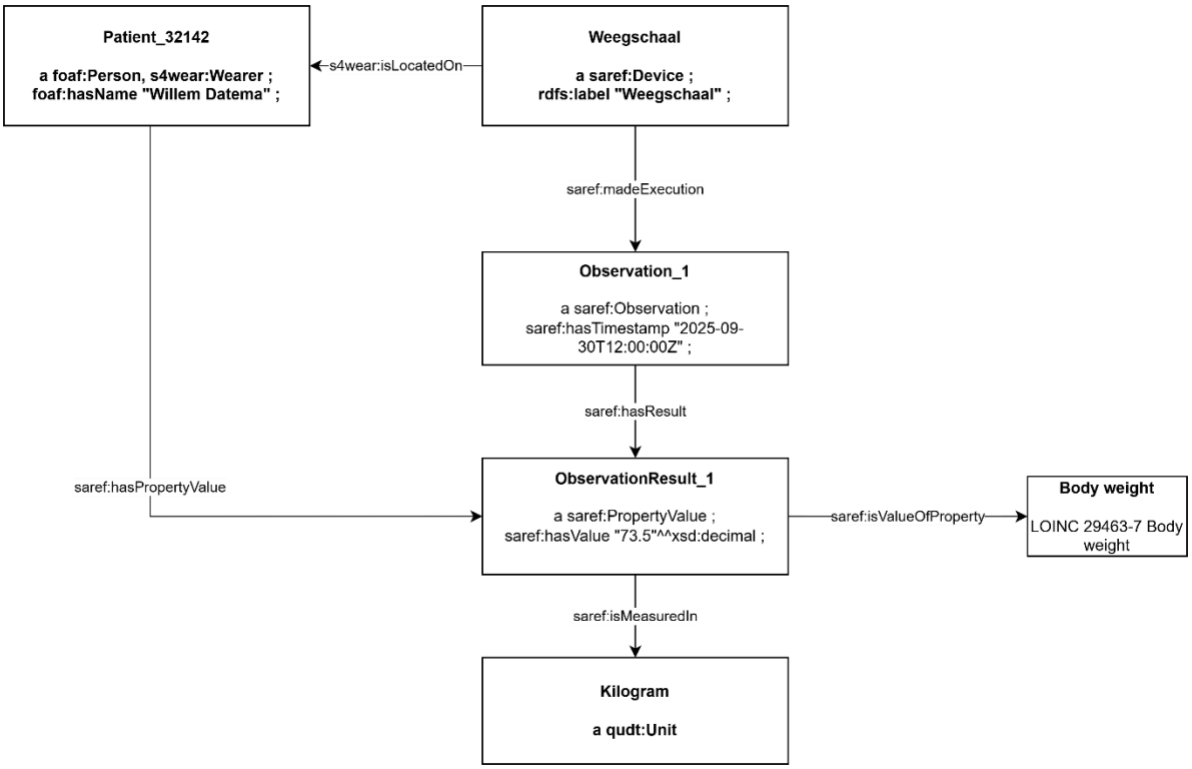

Figure 7 Digital Biomarker Ontology example measurement

The principles of semantic interoperability that are followed by us and by the various semantic standards, ensure that we can reuse each of those in conjunction with each other. Reusability is a key principle of the semantic interoperability paradigm. The generic SAREF and SSN ontologies have deliberately left parts of the model unconstrained, such that domain elements from, in our case, the LOINC and FHIR models can be inserted.

2.5 Discussion & future directions

This section discusses limitations of the current work and future directions we identified while developing the PoC.

First, we showed data interoperability between the SmartScale application and the ZWEM application. This worked because the applications wrote to the same files, using standardized measurements. To achieve true data interoperability, however, some more advanced linked data techniques like SPARQL should be used. SPARQL can be used to query multiple data points, instead of retrieving data from one file like is done in the current implementation. The current implementation requires applications to save the data in the same place, which needs agreements between different app developers. There are some initiatives from the Solid community, like Type Indexes[11], that can indicate in a standardized manner where files are located, but this still requires manual parsing at the moment, as the libraries from Inrupt do not have anything in place for this.

With data interoperability in place, this also opens possibilities for reusing the data for research purposes. This is a topic with attention of the European Commision[12] and is referred to as ‘secondary’ use of data in European law. While the benefits of reusing the data are clear, privacy regulations, requiring consent of the user for each research purpose, make using this data more difficult. This is because the burden on the user should not increase too much. In addition, this requires advanced data anonymization and other privacy measurements, as linking of metadata or outliers in data points could be used to trace an individual even though the data itself is anonymized.

Furthermore, what we also have not tested yet is how the application performs when there are many Solid pods in place. What if, for example we need to read 1000+ pods to compute the statistics? A real-world deployment would require load testing and optimization, for example via simulations.

Another limitation regards the burden the usage of this different technology will place on the users. Will they be willing to adopt this technology if it makes the procedure even just a little more complicated? The Solid technology has various applications for this, and various applications are working towards this, but it may nevertheless inevitably be the case that the data insertion procedure changes.

Finally, currently, the pod provider can read all the data that is generated by the ZWEM application. This is actually useful for demo purposes, but in a real-world setting, we would probably want some client-side encryption in the ZWEM application. This ensures the pod provider cannot read any sensitive user data.

2.5.1 Perceived developer experience with Solid

The application as presented uses two authentication methods. One for the ZWEM application itself (uses credentials stored in the central database) and the Solid Open ID Connect (Solid-OIDC) flow for the Community Solid Server. Ideally, these two login methods

[11] https://solid.github.io/type-indexes/#private-type-index
[12] https://health.ec.europa.eu/ehealth-digital-health-and-care/reuse-health-data_en

would be merged into one Single Sign On (SSO) login method. Most modern applications work with these SSO standards like OIDC and OAuth 2.0. Unfortunately, Community Solid Server (CSS) does not (yet) integrate seamlessly with these standards. This means that solid applications built on top of the Community Solid Server cannot use one login for both the application and the solid access. To enable true SSO between other authentication providers like Keycloak and CSS, middleware or adapters are required to bridge the protocol gaps between the SSO standards and Solid-OIDC, a modified version of OIDC. Future work should implement this middleware, or users could migrate from the community solid server to Inrupt Enterpise Solid Server[13]. This Solid implementation has a Solid OIDC Broker implementation that acts as compatibility layer between decentralized Solid authentication and traditional OIDC provider flows.

Furthermore, improving user interfaces for viewing and editing data directly within Solid pods should be a priority for the Community Solid Server, making decentralized data management intuitive and practical for all participants, regardless of their technical background.

Additionally, the learning curve of the Community Solid Server is quite steep, mainly due to the lack of documentation, which is something the team behind CSS recognizes[14]. What does not help is that there are also other implementations of a Solid server such as the Node Solid Server[15] which is very similar but different. It would help if the community could focus on one implementation on the open-source side, or explain clearly what the strengths and weaknesses of both implementations are. The client libraries of Inrupt, however, are easy to use and make working with authentication and data reading and writing very straightforward.

The community behind Solid consists of many very enthusiastic people who work hard to create demo applications, write standards and use Solid for many different purposes. These people are backed by a more formal structure of the Open Data Institute which streamlines standardization and community efforts. They recently established the W3C Linked Data Storage Working Group which will standardize the Solid ideas on W3C Recommendation level. This is a solid basis for further maturation of the Solid technology.

All in all, we recognize the need to enhance the underlying infrastructure. This includes evaluating the maturity and suitability of alternative (commercial) Solid implementations, which was left out of scope for this research.

[13] https://docs.inrupt.com/ess/latest/services/service-oidc
[14] https://communitysolidserver.github.io/CommunitySolidServer/latest/
[15] https://github.com/nodeSolidServer/node-solid-server

3 Conclusions

This work successfully demonstrated that privacy-friendly data storage is feasible for health collectives using personal data vaults, specifically Solid pods, without sacrificing user experience or the simplicity of the application. The proof-of-concept showed that both personal and aggregated health data can be managed and shared in a decentralized manner, empowering users with full ownership and control over their sensitive information. This work improved data interoperability by adopting semantic standards and ontologies (such as FHIR, LOINC, QUDT, SNOMED CT, SAREF, and SSN), enabling data reuse across applications and domains.
The implementation highlighted the importance of user consent and granular data sharing by demonstrating the ability to opt out of aggregated statistics. The Community Solid Server was functional for demonstration purposes, but was found to be insufficiently mature for real-world deployments. Key limitations identified include the need for single sign-on (SSO) integration, client-side encryption, performance testing at scale and more advanced data interoperability mechanisms such as SPARQL.
Overall, the concept of user-controlled data storage shows significant promise for privacy-preserving health data storage and should be further developed and invested in, particularly in open-source contexts.

3.1 Future directions

To ensure broad adoption, it will be essential to investigate the extent to which users trust Solid technology and to understand their willingness to embrace new procedures, especially if these introduce additional complexity. Usability improvements are also a priority, such as enabling coaches to help participants by entering measurements on their behalf, thereby lowering barriers to participation. For this it is also important to involve actual users.

Another important direction is to facilitate the responsible secondary use of collected data for research purposes. This will involve implementing robust mechanisms for user consent and data anonymization, ensuring that individual privacy is never compromised even as aggregated insights are generated to demonstrate the effectiveness of initiatives like ZWEM. Integration with broader health data infrastructures, such as the European Health Data Space (EHDS), is also on the horizon, which would enable seamless and secure data exchange across national and organizational boundaries.

ICT, Strategy & Policy

www.tno.nl