National Health Care Institute

Final report on the Personal Health Train feasibility assessment

OCTOBER 2018

Inhoud

Management summary 3
Introduction 4

1 About the Personal Health Train (PHT) 5
1.1 Relationship with FAIR Data principles 5
1.2 Train that passes data stations 5
1.3 Advantages 5
1.4 Development and research 5

2 Feasibility test based on a case study 6
2.1 Intra-arterial thrombectomy case study 6
2.2 Possibilities with the PHT 6
2.3 Using the PHT in a practical situation 6

3 Parts of the simulation environment 7
3.1 FAIR Data Point 7
3.2 FAIR Data Station 7
3.3 Train 7
3.4 Rails 8
3.5 Dashboard 8

4 Operation in practice 9

5 Implementation: agreements, impact and privacy 10
5.1 Impact on the organisation 10
5.2 Technical standards 10
5.3 Full set of agreements 10
5.4 Train can operate on small-scale networks 10

6 Legal points of attention for further PHT development 11
6.1 Who has ultimate responsibility for the data? 11
6.2 Example scenario 11
6.3 Legal analysis 11

7 Conclusions 12
7.1 General aspects 12
7.2 Technical background 12
7.3 Privacy and safeguards 12
7.4 Transparency 12
7.5 Implementation 13
7.6 FAIR Data Services or PHT 13
7.7 Legal conclusion 13

8 Further steps 14

9 Pels Rijcken Advisory Report 15

Management summary

This is the final report of the FAIR Data part of the feasibility assessment on FAIR Data & the Personal Health Train (PHT). In the assessment, *Zorginstituut Nederland* applied the FAIR Data principles to its own data and then implemented the PHT in a practical situation. The activities were carried out in accordance with agreements between the Minister of Health, Welfare and Sport (VWS) and the *Zorginstituut*.

This final report describes what the *Zorginstituut* achieved by using a PHT. This final report gives an overview of the results and findings of the second phase: PHT. A separate final report is available on the first phase (FAIR Data).

The most important concept of the PHT is that data are not taken to be analysed, but analysis is taken to the data. The data remain at source – in accordance with the outcome objectives of the Information Debate – thus creating an opportunity to carry out complex, self-learning algorithms at source. Furthermore, PHT interprets the privacy-by-design principle by using detailed personal data itself in analyses, while only aggregated data are included in, for instance, the conclusions that go back to the user. These advantages became apparent in the feasibility assessment; they facilitate reams of data applications that are currently not possible.

Another important finding is that we envisage PHT in the light of a broader context of FAIR Data Services. The most suitable form of PHT has to be determined depending on the situation and the organisation that a station offers for the PHT, partly because some forms of PHT implementation are technically and organisationally more complex.

More research and harmonisation is required for a broader implementation of the PHT. This involves questions about technology, privacy and organisational aspects that first need to be answered. The PHT could play a role in the near future in smaller networks, in which a large degree of trust already exists between the parties. Section 8 makes a few suggestions for further research.

Introduction

One of the *Zorginstituut*'s tasks is to interpret and research new developments in information management, with the objective of taking the most promising applications a step further and ultimately improving health care.

The amount of health care data is growing rapidly; data not only of care providers and researchers, but also of care consumers. These are important data that can help to improve health care.

But how can they be accessed? And who can put them to good use? The *Zorginstituut* examined two concepts that can help: FAIR Data and the Personal Health Train (PHT).

Between October 2017 and May 2018, the *Zorginstituut* worked on the FAIR Data & Personal Health Train feasibility assessment. These concepts originated in the academic world and have received a lot of attention in recent months. They are mainly studied and used in relation to research and development.

By means of this feasibility assessment, *Zorginstituut Nederland* studied FAIR Data and PHT in more depth and used it in an operational environment within health care. Implementing the assessment involved two phases.

The focus of the first phase was on using the principles of FAIR Data and exploring its implementation in health care. In the second phase, a simulation PHT was realised.

This final report gives an overview of the results and findings of the second phase: PHT. A separate final report is available on the first phase, FAIR Data.

Contact

Questions about the FAIR Data & Personal Health Train can be sent to:

Zorginstituut Nederland
Wouter Franke, project manager, wfranke@zinl.nl / 06 17 73 09 39
Team Information standards (Information management department)

info@istandaarden.nl
www.istandaarden.nl

Office address
Willem Dudokhof 1
1112 ZA Diemen

Postal address
Postbus 320
1110 AH Diemen

1 About the Personal Health Train (PHT)

In this section we explain the most important PHT concept: that data are not taken to be analysed, but analysis is taken to the data. The data remain at source, in accordance with the outcome objectives of the Information Debate.

The Personal Health Train is a continuation of the FAIR Data principles. These principles describe a meticulous and measurable set of qualities on which every good data publication must be based [1]. These principles and their elaboration help to provide an answer to the growing challenges set by data, such as their abundance, the diversity of standards, interoperability and fragmentation.

1.1 Relationship with FAIR Data principles

FAIR Data-principles are general principles that apply to good data management. In fact, FAIR Data are about the data source. The Personal Health Train is a concept involving the use of these data, e.g., for analysis purposes.
In general, regular data analysis starts with collecting data in a single location. This is where data are cleaned up, checked for completeness and – when the data come from different sources – translated according to a common definition.

1.2 Train that passes data stations

The most important characteristic and starting point of the Personal Health Train is that data are not taken to be analysed, as the analysis is taken to the data. It is similar to a train that passes by: the data remain at source and the owner allows them to be used by the passing train for the purpose of analysis. When data from different sources are used, the train passes different stations to analyse the data. During the analysis the data remain at source; the train takes only the conclusions back to the analyst.

1.3 Advantages

For privacy, the PHT is a better concept than collecting all data in one location. Furthermore, it avoids the duplication of data, which involves risks regarding the data's accuracy and completeness. A third advantage of the PHT is that the data-owner retains much more control over the data because he determines the rules for his 'station'.

One important aspect is that more data will become available for analysis with the PHT once there is sufficient support for using the train. An enormous *potential* advantage is, therefore, that the more and better control data-owners have, the more data will become available. The assumption is that this self-reinforcing effect will eventually occur.

1.4 Development and research

The PHT concept was developed by the Dutch Techcentre for Life Sciences (DTL)[2], in collaboration with MAASTRO and the LUMC. Sub-topics of the concept have been consistently studied in a number of research trajectories. MAASTRO[3] studied the validity of the self-learning algorithms distributed[4], after which an initial test run[5] took place with the PHT[6]. The *Zorginstituut* made use of this acquired knowledge in a simulation of the PHT in an operational environment.

1 https://www.nature.com/articles/sdata201618
2 https://www.dtls.nl/fair-data/personal-health-train/
3 https://www.sciencedirect.com/science/article/pii/S2405630816300271
4 https://www.sciencedirect.com/science/article/pii/S2405630816300271
5 https://www.ncbi.nlm.nih.gov/pubmed/29678027
6 https://www.ncbi.nlm.nih.gov/pubmed/29678027

2 Feasibility test based on a case study

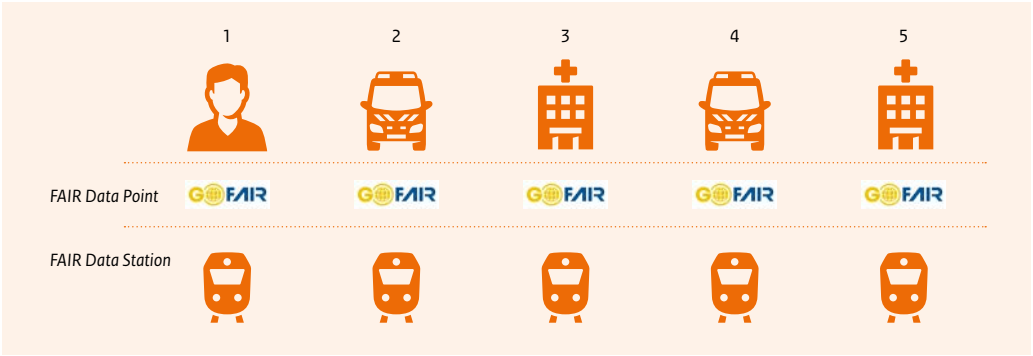
For the feasibility assessment the Zorginstituut realised an environment with various data stations where a PHT could carry out its analyses. To be able to check its operation, a scenario was drawn up in which the times and sequence of steps in the process play an important role in supplying good care. The scenario involves a patient who is undergoing an inter-arterial thrombectomy (IAT), a relatively new form of treatment.

2.1 Intra-arterial thrombectomy case study

Intra-arterial thrombectomy (IAT) is treatment for a brain infraction whereby the surgeon removes the blood clot through a catheter inserted via the groin. This is popularly known as 'percutaneous angioplasty in the brain'. It is a highly complex treatment that twelve authorised centres have been designated to perform in the Netherlands. IAT must take place within a number of hours after the infarction occurred. This means that the time between the various steps in processing a patient is of crucial importance.

2.2 Possibilities with the PHT

Patients, scientific associations, researchers and quality organisations may be interested in the processing times of IAT treatments. The PHT offers possibilities for collecting this information retrospectively.



2.3 Using the PHT in a practical situation

Which steps did the patient go through and how long did each of these steps take? Where could we realise improvements in the process?

To ask – and answer – this question with the help of the PHT in this case study, we followed the processes that the patient went through (see the diagram) and added PHT elements.

- The patient suffers a brain infarction. He calls the emergency number (112) to be rushed to hospital.
- The ambulance (Friesland services) picks the patient up and takes him to a hospital in Leeuwarden. The ambulance service publishes relevant information from its own registration system on a FAIR Data Point.
- Doctors in the regional hospital in Leeuwarden diagnose that the patient has had a brain infarction and he receives intravenous thrombolysis (IVT). In addition, the hospital suspects that the patient is eligible for IAT treatment and a second ambulance is organised. The hospital also publishes relevant information from its own registration system on a FAIR Data Point.
- The second ambulance (Groningen services) takes the patient from the hospital to Groningen UMC (UMCG), a hospital authorised to perform IAT. The ambulance service publishes the relevant information from its own registration system on a FAIR Data Point.
- The UMCG carries out the IAT treatment and publishes only its own relevant information on a FAIR Data Point.

In order to obtain the information on processing times from this chain, the PHT starts its journey, driven by the patient who wants this information. The train passes various stations, analyses the data located there, determines its route and finally sends the results of the analysis back to the patient.
In the actual situation, which we describe in the next section, it is all about the processing time indicators.

3 Parts of the simulation environment

In the second phase of the feasibility assessment, the scenario was elaborated upon in a simulation environment. A number of aspects were realised there.

3.1 FAIR Data Point

The location where the data-owner stores his data. Four data points with relevant data were realised for the simulation: one for each organisation in the case study. An ontology and examples of data were drawn up for each data point. An ontology is a description of a reality, or representation of knowledge, that computers are capable of interpreting.

In general, organisations make a selection of operational data available on their data point and they equip the data with an ontology in accordance with the FAIR Data principles. The data are then placed on the data point.

In the feasibility assessment we used the open source editions of OpenLink Virtuoso for the data point. Virtuoso is one of the most frequently used databases for storing so-called *linked data*-triples (object – predicate – subject). These databases are known as *triple-stores or RDF-stores.*

3.2 FAIR Data Station

The station that offers services for accessing the data. During implementation we concluded that there are various types of services and, therefore, numerous technical implementations. Depending on the intended use (application), the service has to be more or less secure, and depending on the complexity of the analysis, the service has to offer more or less possibilities.

We realised three types of services that a data station can offer.

1. **Sparql-endpoint** for questions where privacy is less important, as with *open data*, but which do offer an analyst many possibilities for defining his own questions. This service is relatively simple to implement and offers an opportunity of developing new applications for data. In the case study, each station has a Sparql-endpoint.

2. **RESTful-interface** to be able to adequately guarantee privacy and define questions in advance. This service offers an analyst little flexibility, is more complex to implement than a Sparql-endpoint, but does offer a lot of certainty regarding security and privacy. In the case study, the two ambulance stations offered services via this interface, alongside the Sparql-endpoint.

3. **Docker container** In this environment, the analyst can produce an algorithm for the data-owner. This service comes closest to the metaphorical train. In the feasibility assessment, a *docker container* was chosen as location for the algorithms because this technology makes the software independent of the environment in which it operates. Good guarantees are possible for privacy and an analyst also has many possibilities for designing the analysis. From a technical perspective, implementing this service is more complex than the first two services. Furthermore, agreements between users of the services are necessary in order to create the right guarantees in the field of privacy, technology and organisation. The stations of both hospitals offered this interface for implementing the case study.

3.3 Train

The train forms the logics of the analysis. The form the train takes varies according to the type of service. The train may take the form of a query, a request by an interface or a fully programmed and self-learning algorithm. At the *docker container*, the train always first sends a request to the data-owner, asking for permission to carry out the analysis on site.

3.4 Rails

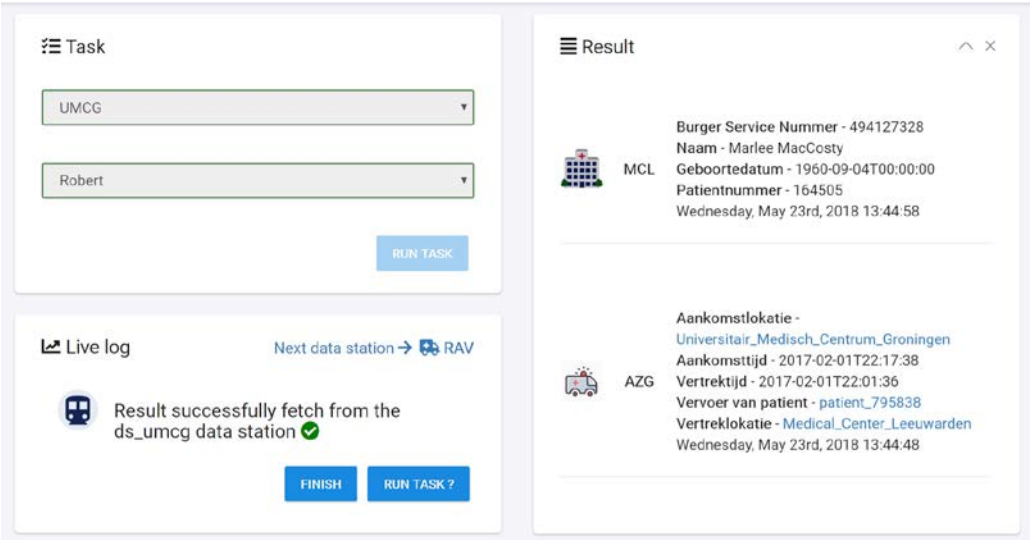
The 'rails' are the agreements, guarantees and interfaces of the PHT. The rails are not physical infrastructures like VPN-connections. Organisations can easily get involved, with existing resources, without needing any central infrastructure.

The rails are agreements about the type of services provided by a station, about which logging and audit trails we want to arrange communally, what authentication we will use, etc. Ultimately, the agreements can take on the form of a set of agreements or something similar for establishing and safeguarding guarantees.

Ultimately, a PHT does require a certain degree of trust. We still have to discover exactly which guarantees and criteria will be needed to get this trust. In our opinion, leaders in the field of PHT are needed to set an example and determine the direction to be taken.

3.5 Dashboard

This is where a monitor is used to track the progress of the 'train'. The dashboard (see illustration) is where the analysis starts and the results are displayed.



4 Operation in practice

Using parts from the simulation environment, in a number of steps, the user case tested how the PHT worked in practice.

1. The analysis started from the dashboard by sending a request to the UMCG to supply an analysis. The analysis could also have been started from one of the other FAIR Data Stations, but in the feasibility assessment we opted to start the analysis at the last place visited by the patient. The analysis request also states who wants the analysis to be done and where to download the container that can carry out the analysis.

2. The UMCG can automatically assess this request and, after approving it, retrieve the *docker container* with the analysis. The first step of the analysis is carried out within the UMCG's environment. The analysis also includes the treatment times (IAT) of the UMCG and the information that the patient arrived by ambulance.

3. Based on this information, the container asks the data station of Groningen's ambulance service for the ambulance times and the name of the transport operator. Based on this information, the container sends a request to the MCL's data station to collect a *docker container* for the analysis. This request also states who wants the analysis and where to download the container that can carry out the analysis.

4. The MCL can assess this request automatically and, after approving it, retrieve the *docker container* with the analysis. Again, part of the analysis is carried out within the MCL's environment. The container receives the treatment times (IVT) of the MCL and the information that the patient arrived by ambulance.

5. Based on this information, the container asks the data station of Friesland's ambulance service for the ambulance times and the name of the transport operator.

6. All partial reports are sent back to the dashboard where the steps in the process have now all been registered.

This PHT process can be used for a group of patients as well as for individual patients.

5 Implementation: agreements, impact and privacy

To implement a PHT successfully it is essential that agreements are made with all the parties involved in the train. In theory, even organisations in a small-scale network can start implementing a PHT. The rails form the basis for every network that wants to use the PHT. In other words: good agreements.

The rails are formed by the joint agreements for allowing the PHT to travel successfully and safely around in a network. In which fields and to what extent are agreements needed between parties in the network? What has most impact? And what do we do about privacy?

5.1 Impact on the organisation

A FAIR Data Service in the form a *docker container* has most impact on organisations in the network. This is because an external party places software in the organisation's own environment and gets access and rights to use software on a dataset. This infringes on measures the organisation has taken to protect data. Only confidence, anchored in clear agreements, can make this privacy matter manageable.

5.2 Technical standards

The rails are essential to allow the PHT to function properly. This requires agreements on the technical standards of a FAIR Data Station. For example: which self-service interface do you use to roll out and operate trains on a FAIR Data Station? Such interfaces are designed as much as possible to support authorised users (of data) automatically. A task manager was implemented for this in the feasibility assessment, though there are various alternatives.

5.3 Full set of agreements

Agreements are also needed in the field of workflow and database-interfaces. Expectations are that these agreements will lead to a complete set of agreements on the development, roll-out and management of FAIR Data Stations and the accompanying FAIR Data Services. For the technical elaboration of this set of agreements, possibilities include cloud suppliers such as Amazon, Google, Microsoft, IBM and various smaller parties in the market. They could offer FAIR Data Services that allow individual care providers to realise a FAIR Data Station. Also expected is that the PHT will be linked to a price model, to cover the costs of using the processor and infrastructure.

5.4 Train can operate on small-scale networks

Other solutions are possible for small-scale networks that already enjoy a greater degree of confidence. In its simplest form, bilateral agreements will be made on the roll-out and operation of a PHT in the environment of a data source. These types of networks could be comprised of a group of hospitals, a regional group of nursing homes, research institutes, a number of suppliers of personal health environments (PGOs) or various public organisations. These types of networks could already start implementing a PHT.

6 Legal points of attention for further PHT development

To look at the PHT from a legal perspective, we cooperated with experts to draw up an initial analysis with a view to privacy aspects. The analysis provides a full overview of legal questions about the PHT, supplies advice and draws conclusions.

The document, written by Nina Bontje, lawyer at Pels Rijcken, contains mainly points of attention for the further development of the Personal Health Train.

6.1 Who has ultimate responsibility for the data?

When processing personal data, the most important question is: who is ultimately responsible for the data? There is no problem, as long as the analysis carried out by the PHT does not use personal data. When it does, then agreements are needed. We explain these with the help of an example scenario.

6.2 Example scenario

Take the case of a researcher in a hospital who wants to produce an analysis. He will be interested in aggregate data, but during the analysis he will also need access to personal data. In accordance with the PHT, privacy-by-design is taken into account, so only the analysis, i.e., the *docker container*, is given access to the personal data. The results that the researcher eventually sees are only totals, without any personal data. The question here is: who – according to the AVG legislation – is responsible for the data and who is the data processor?

6.3 Legal analysis

Following analysis from a legal perspective, in this scenario it seems that both parties bear joint responsibility. However, it is not immediately clear how the researcher fulfils the requirements. After all, he knows only the type of data that was used for the analysis, but is not acquainted with the actual data. Thus, a researcher can only organise a person's request for insight into data in collaboration with the data-owner. To do this, the PHT system needs agreements.

The full title of Pels Rijcken's report is *Points of attention for further development – Privacy Aspects of the Personal Health Train*. This report is included in full as section 9.

7 Conclusions

The origin of the FAIR Data & Personal Health Train feasibility assessment lies in the workspace of research and development. Zorginstituut Nederland has examined its applicability in a broad sense, in more operational environments – such as process and quality information. In this section we draw conclusions based on the assessment.

7.1 General aspects

An important aspect is that more data will become available for analysis with the PHT once there is sufficient support for using the train. Although the PHT currently being developed has not come as far as FAIR Data, the *Zorginstituut* is convinced of the opportunities that the PHT offers for data-analysis and the necessary interoperability that is currently not – or barely – being realised.

It is important to look upon the PHT as a part of FAIR Data Services, as it is not limited to a single PHT. The PHT comes in various formats, or services, depending on the characteristics of the data-owner and the complexity of the analysis. Implementing a *docker container* for a fairly simple question is a form of 'over-engineering'. For complex algorithms, e.g., artificial intelligence (AI), machine learning and intelligent agents, this form of implementation has a real added value in comparison with current methods.

For less complex questions, other FAIR Data Services – i.e. other PHT formats, such as a Sparql-endpoint or a RESTful-interface – are highly useful technologies. When taking a broader look at the FAIR Data Services, the train metaphor falters, though the strength of the concept of 'taking the analysis to the data' remains undiminished.

7.2 Technical background

As mentioned earlier, there are various PHT formats. This means various implementations are possible. For instance, a Sparql-endpoint as implementation works well for *open data*, and a RESTful-interface works well for an application where an interface's input and output are defined precisely and can therefore be optimised.

Implementing the PHT with the help of *docker containers* works well for complex analyses when there are large data variations, for example, due to many different locations or due to a multitude of definitions. In the feasibility assessment, we easily made effective applications that display the power of these services.

Realising a PHT does not demand an enormous central infrastructure. The starting point is to offer a number of services to make data available. These are the FAIR Data Points and the FAIR Data Stations. These services are realised by organisations when they want to share and use data. The rails that are needed to allow these services to communicate with one another are mainly the agreements that are needed on technical interfaces, implementation of authorisation and authentication, and guarantees such as audit trails.

7.3 Privacy and safeguards

The PHT has aspects that make privacy-by-design possible. Data-owners retain control of their data and the safeguards in the system can secure the balance between flexibility and privacy. For instance, you can carry out analyses on combined data that were collected in more than one location. All that the analyst receives, however, are the results of the analysis. The underlying data remain with their owner. As mentioned earlier, these privacy guarantees – this confidence – require agreements that go further than the technology used. The question is to what extent such agreements are scalable, partly due to the potentially international nature of the PHT.

7.4 Transparency

FAIR Data operates based on transparency. The principles are mainly about 1) stating what data you have, 2) equipping them with rich definitions and 3) clearly indicating how a person can access the data. On the other hand, a degree of trust plays a role. The PHT involves a request mechanism, whereby the analyst

asks the data-owner for permission to carry out an analysis. Various guarantees for transparency can be introduced into this process. And when a PHT is used to carry out an analysis at data-source, the data-owner can clearly see which data are being used. This contributes to a more transparent information system.

7.5 Implementation

Implementing the PHT in the feasibility assessment took a few months. The time invested was mainly spent on realising a number of components for which few example implementations were available. A number of architectural questions have still not been answered – though they will have to be answered to be able to implement the PHT structurally. We made choices in the feasibility assessment, such as using the *docker containers* or the metadata that are used for a request. These choices will have to be discussed when developing the PHT further. Furthermore, implementation generally takes place in networks, just as the case study took place in a network of sequenced care. Harmonisation will always be necessary for such implementations.

7.6 FAIR Data Services or PHT

The train metaphor falters when the concept is developed in more detail. The moment will come when using the PHT as an intelligent agent would lead to the metaphor of trains that are stationary. At the same time, the PHT offers a multitude of opportunities for PGOs to carry out data analyses and to make patient data available, for instance, for research.

But a PHT application could also be at a different level than the personal client level. For this reason, during the feasibility assessment we increasingly started using the term FAIR Data Services instead of PHT. This gives a broader picture of the potential for a network of organisations linked by ontologies. A network in which microservices are offered to make interpretations possible and new valuable services can be realised. It is important to see the PHT as part of the larger entity of FAIR Data Services.

7.7 Legal conclusion

The conclusion from the report of Pels Rijcken & Droogleever Fortuijn, lawyers and civil notaries, 'Points of attention for further development - Privacy Aspects of the Personal Health Train', reads:

"(...) Our conclusion is that no red flags are evident that can be expected to impede further development of the PHT. There are orange flags that the Zorginstituut should take into account during its further development.

We identified these orange flags based on ten in-depth questions. We consistently indicated an initial possible solution. Naturally, we are prepared to elaborate on these possible solutions, either before or after the various uses cases of the PHT have been elaborated upon."

8 Further steps

The feasibility assessment has given insight into the working and implementation of the PHT. As mentioned previously, the Zorginstituut is positive about these developments.
The scalability of the PHT in particular has not yet been sufficiently elaborated upon, specifically in implementing FAIR Data Services. Furthermore, there are various architectural and legal questions:

- Feasibility in practice and monitoring operational deployability. Based on the feasibility assessment, we are confident that the PHT can be deployed operationally, partly because the PHT largely uses existing technology. This could be tested by including – and thus operationalising – PHT implementation in the simulation model of the iWlz Action Programme.

- Unanswered legal questions, with the main question who is the data processor and who is responsible for the data in PHT situations. This will help provide input for the guarantees (agreements) that are needed for the PHT. These were subjected to an initial exploration together with Pels Rijcken's lawyers. The conclusion of their advice report is provided in section 9.

- Unanswered technical questions that focus mainly on safeguarding data within the various services of a data station. An example is the degree to which RDF-stores can regulate authorisation at an elementary data level. The feasibility assessment generates sufficient confidence that this is possible, though the best way to realise this is a part of follow-up research.

- Making the necessary agreements to realise an Internet of FAIR Data Services within health care. The rails, or set of agreements for the PHT, must take such a form that several parties can implement the PHT. This will mean a coordinating role for government, until the mindset and technology have become so commonplace that the parties do it themselves.

9 Pels Rijcken Advisory Report

Privacyaspecten van de Personal Health Train

Aandachtspunten voor de verdere ontwikkeling

N.N. (Nina) Bontje
30 juni 2018