UNIVERSITY
OF TWENTE.

Faculty of Electrical Engineering, Mathematics & Computer Science

Automated machine-readable data access agreements by applying ODRL to a FAIR Data Train

Siem Veltmaat
S2172143
s.p.veltmaat@student.utwente.nl

MSc Business Information Technology
Master Thesis
September 2024

Supervisors:
dr. L.O. Bonino Da Silva Santos
dr. R. Guizzardi - Silva Souza
dr. L. Ferreira Pires

Abstract

Nowadays, daily data generation is at an all-time high and is predicted to further grow in the future. Therefore, the demand for efficient, automated, and secure data access agreements is becoming increasingly crucial. This thesis investigates the potential of using the Open Digital Rights Language (ODRL) to create machine-readable data access agreements within the FAIR Data Train (FDT) framework. The main objective is to explore how ODRL can facilitate data access authorisation in federated analysis platforms like the FDT, thereby enhancing the efficiency of data sharing while maintaining data ownership and adhering to privacy regulations.

The research follows the Design Science methodology. Which involves the problem investigation, solution design, and validation phases. Initially, a comprehensive literature review on Rights Expression Languages (RELs) was conducted to assess the suitability of ODRL for this application. The design phase involved creating scenarios demonstrating various aspects of ODRL defining data access agreements, which were then validated through stakeholder surveys within a real-world FDT use case.

The findings indicate that ODRL can effectively support the creation of standardised, automated data access agreements. The developed scenarios and corresponding Resource Description Framework (RDF) agreements provide a robust basis for matching access requests with access policies. This matching process is critical for automating data access authorisations, significantly reducing the time and manual effort traditionally required.

In conclusion, the research confirms that applying ODRL within the FDT framework offers a viable solution for automating data access agreements. This advancement streamlines the data sharing process while upholding essential privacy and ownership standards.

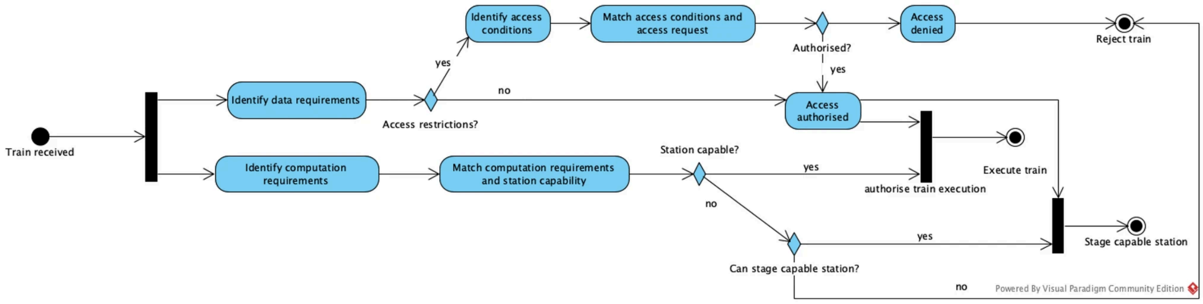
Keywords: *FAIR Data Train (FDT), Personal Health Train (PHT), Rights Expression Language(REL), Open Digital Rights Language (ODRL).*

Content

1. Introduction

Nowadays, data generation is at an all-time high, with Statista[1] predicting that this will increasingly grow in the near future. Because of the mass generation of data, it becomes even more important that processes involving data are optimised and efficient to reduce future bottlenecks. Together with the upcoming rise of industry 5.0 and the increase of machinery in business processes, it becomes more important that business processes get automated. To support the goal of automation in businesses, data should be machine interpretable, so that machines can fulfil their tasks in a correct manner, and semantics play a big role in this [2]. Machine interpretable data can be produced by following the principles of making data Findable, Accessible, Interoperable and Reusable (FAIR) [3].

In most areas, such as in healthcare, data is often privacy-sensitive and thus a strict process is involved to share data. Organisations have to comply with strict regulations like the GDPR for example [4]. In most cases, the access conditions are not explicitly stated and, much less, available in a machine-actionable format. The same holds for data access requests. However, recently the concept of the Personal Health Train (PHT) was introduced [5]. The PHT is based on the FAIR data principles and in short attempts to bring algorithms (trains) to the data (stations) instead of the other way around, in a similar manner to how a train in the real world wants to deliver passengers to train stations. This way, the organisations still have control over the data and can check the purpose and intended result of an algorithm. The PHT is based on medical data, but this concept can be applied to any kind of data. Therefore, a more suitable name would then be a FAIR Data Train (FDT) and this will be the terminology applied in the rest of this thesis [6].



Train received
Identify data requirements
Access restrictions?
yes
no
Identify access conditions
Match access conditions and access request
Authorised?
yes
Access denied
Reject train
Access authorised
Identify computation requirements
Match computation requirements and station capability
Station capable?
yes
no
authorise train execution
Execute train
Can stage capable station?
yes
no
Stage capable station
Powered By Visual Paradigm Community Edition

Figure 1: High level steps of the train evaluation process [5].

Figure 1 shows the train evaluation process of the FDT. With its foundation on the FAIR principles, the FDT makes extensive use of metadata and machine-actionable declarations. As seen in Figure 1, whenever a train tries to enter a station, a check is performed if the station has access conditions in place. Whenever this is the case, it means that the FDT should have an access request specified in its metadata described in some form of a Rights Expression Language (REL). RELs are a means of expressing the rights of a party to a certain asset [1]. The data station has data access conditions also described in a matching REL. When the train attempts to enter a station, the data access request should be compared with the data access conditions to check whether they match. Based on the result of the matching process, a train access to the data is either granted or denied and, if granted, is executed at the station.

[1] https://www.statista.com/statistics/871513/worldwide-data-created/

Currently, the process of requesting access to data often contains manual actions [32]. An ethics committee is often involved to consider the data access request and grant or deny access to the data for researchers [8][9]. This manual process is time consuming and could lead to errors. Therefore, there is a need for automation to help researchers obtain scientific data in a faster manner.

The Open Digital Rights Language (ODRL) is the W3C recommendation for RELs [7]. As stated by [7]: “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. The ODRL Information Model describes the underlying concepts, entities, and relationships that form the foundational basis for the semantics of the ODRL policies.”

Figure 2 shows the ODRL information model, which is entirely semantic-based. This means that the policies, rules, constraints, etc. are machine interpretable and should thus be machine actionable. Furthermore, being a semantic model, ODRL enables the FAIR principles of being interoperable and reusable.

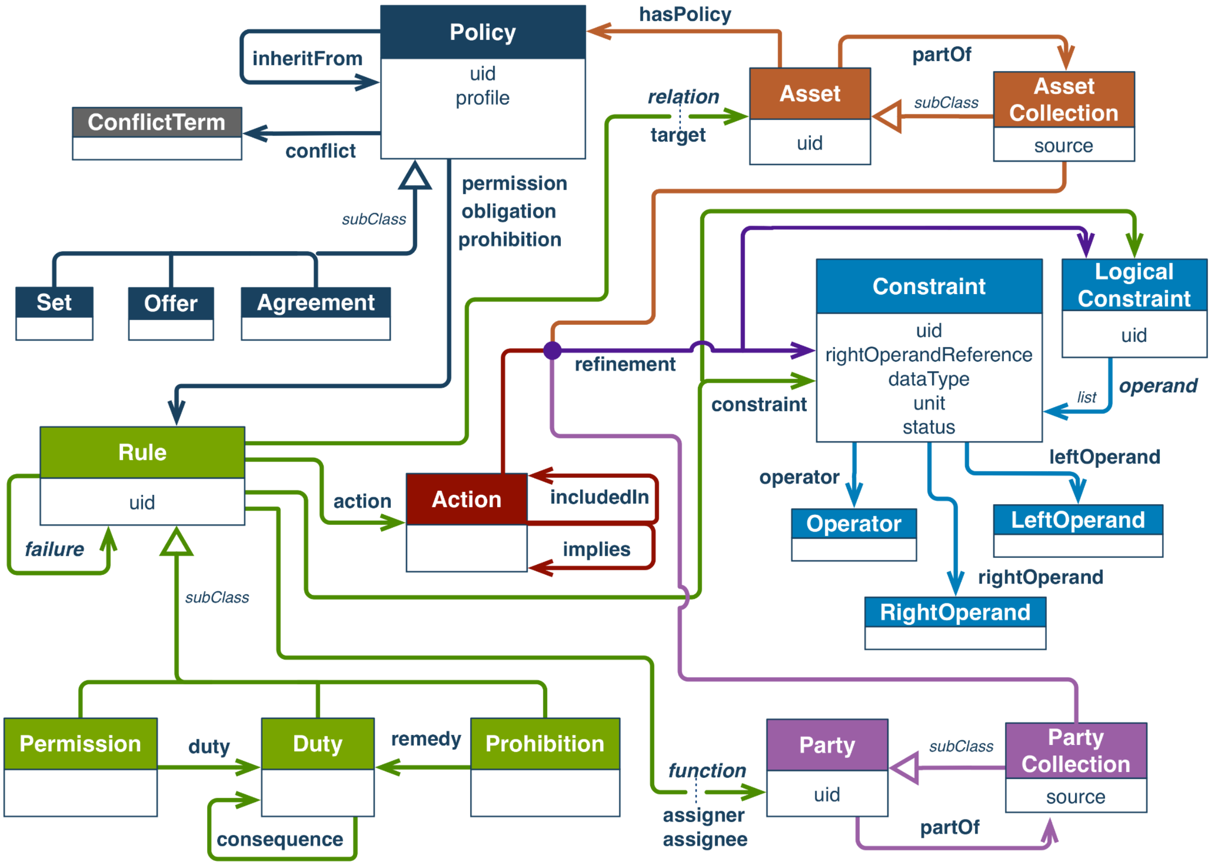


Figure 2: ODRL information model [7]

The main characteristics of ODRL are a policy, an asset, parties, actions, rules and constraints. An asset refers to any type of digital entity, which can be a media file, dataset, scientific article, etc. An asset has a policy, which is usually determined by a party that owns the asset. The policy contains rules and constraints that have to be followed to perform an action on an asset. All of these elements are further explained in Chapter 4.

Furthermore, one of ODRL's key strengths is its extensibility. ODRL can easily be combined with ontologies and vocabularies to further specify conditions, assets etc. in a semantic manner. Moreover, ODRL specifications can be represented in a JSON-Linked Data (JSON-LD) or Resource Descriptive Framework (RDF) Turtle (TTL) format, due to its linked data nature. ODRL also provides the users with the ability to define data access conditions as well as a data access request, a feature that is unavailable in most other RELs. ODRL suits the purpose of describing the access policy of an asset and describing an access request of a FDT, due to its semantic richness and interoperability with ontologies and vocabularies. If the rules and constraints in an access request comply with the rules and constraints in an access policy, an ODRL agreement can be made and sent to both parties.

The aim of this thesis is to evaluate the use of ODRL as a mechanism to describe data access requests and data access conditions in the FDT. The aforementioned strengths of ODRL are part of the reason why this REL has been selected for this research. A comparison between RELs can be found in Chapter 2, including a more detailed description of why ODRL has been chosen.

1.1 Research questions

Currently, a common procedure to get access to private data is to fill out a form stating the purposes of the data collection [32]. Checking these access requests is often still a manual process. The process often goes as follows: first, a person who wants access to data has to fill out a form stating the purposes of the data collection, after which it is manually checked, often by some sort of ethics committee or something similar. If the purpose of using the data is ethical and in-line with the organisation's expectations, access is granted [8][9]. Figure 3 shows the current common process of requesting access to data. In this current process, getting access to the data is too time-consuming to make applying an FDT feasible, since multiple requests have to be sent to each organisation that owns the data required by the train. These requests will then have to be manually reviewed before access could be granted.

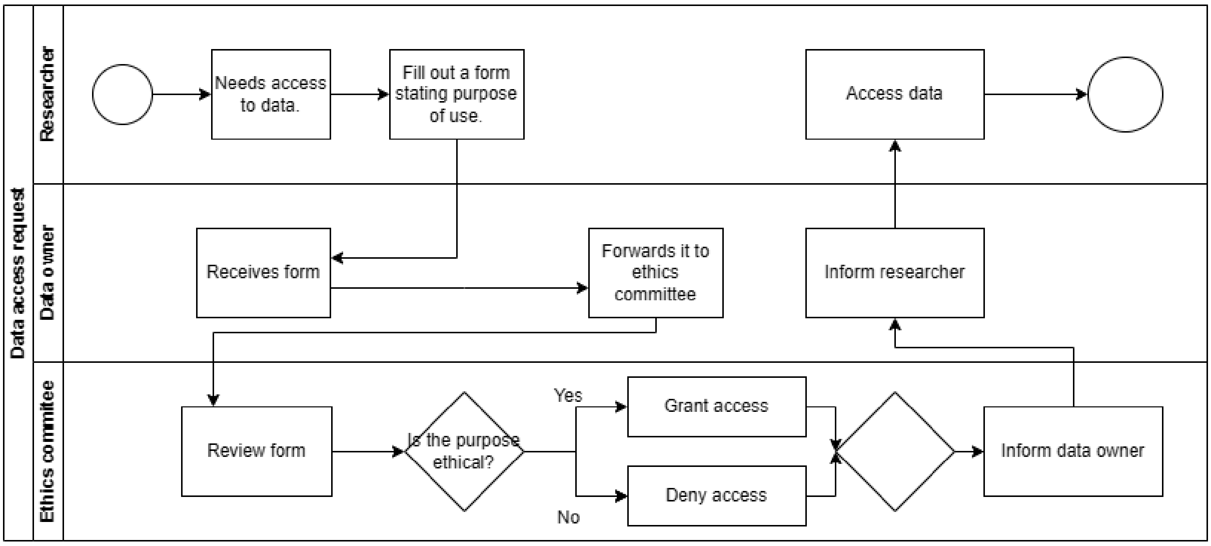


Figure 3: Current process of requesting access to data.

Currently this is already a bottleneck and it will only increase in the future, where automation will be key for businesses. Therefore, the efficiency and speed of this process should be increased through data access automation. In [5], the authors focused on presenting the main concepts of the PHT/FDT without details on how some processes, such as data access request and data access authorization would work. Therefore, an algorithm should be developed to automate the matching process of access requests and access conditions, ultimately granting/denying access of a FDT to a data station. The algorithm should create an ODRL agreement based on the access request and access conditions. Currently, there is no method to create agreements based on the access conditions and access requests. Hence, the main research question of this thesis is:

RQ: How can ODRL be used to support data access authorization in federated analysis platforms such as the FDT?

This main research question can be divided into three sub-questions:

RQ1: Which ODRL elements should be included in a standardised machine-readable access request?
RQ2: What are the desired outcomes of a matching algorithm based on various access policies and requests?
RQ3: What are the steps an algorithm needs to take to create agreements based on offers and requests?

Research question 1 and 2 are answered in Chapter 4. Furthermore, the third research question is answered in Chapter 5. Lastly, the main research question is answered in Chapter 7.

1.2 Methodology

Design Science [21] is the chosen methodology that is applied in this work. The Design Cycle is part of the methodology and is shown in Figure 4. The Design Cycle is part of the Engineering Cycle, the Engineering Cycle consists of four phases: the implementation evaluation/ problem investigation phase, the treatment design phase, the treatment validation phase and the treatment implementation phase. The Design Cycle consists of the first three phases of the Engineering Cycle. The main purpose of this cycle is to identify a problem and design a suitable solution for this problem in the form of an artefact, following the three phases consecutively. The cycle usually starts off at the implementation evaluation/ problem investigation phase. All of the phases are covered in Chapter 3, except for the treatment implementation, because this is just a matter of applying the artefact until the implementation evaluation / problem investigation phase comes around again.

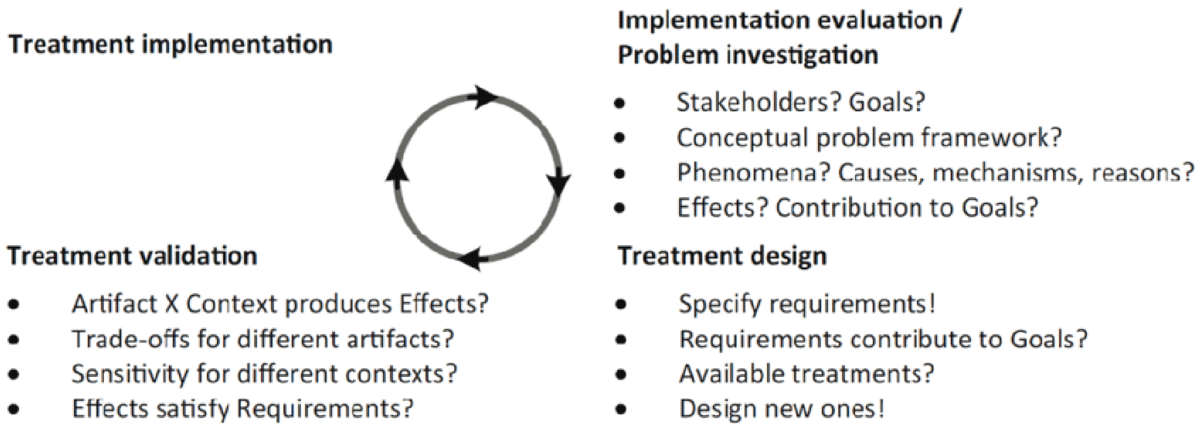

Treatment implementation
Implementation evaluation / Problem investigation
Stakeholders? Goals?
Conceptual problem framework?
Phenomena? Causes, mechanisms, reasons?
Effects? Contribution to Goals?
Treatment validation
Artifact X Context produces Effects?
Trade-offs for different artifacts?
Sensitivity for different contexts?
Effects satisfy Requirements?
Treatment design
Specify requirements!
Requirements contribute to Goals?
Available treatments?
Design new ones!

Figure 4: The Engineering Cycle (Wieringa, 2014) [21].

1.3 Structure

This thesis has the following structure: Chapter 2 presents a literature review that discusses related work and other RELs. The choice for ODRL is also further explained in this chapter. Chapter 3 discusses the phases of the Design Cycle. In this chapter, stakeholders and their goals are identified, requirements are defined for the artefacts that have been designed and the validation method is discussed. Chapter 4 presents the results of the research. This chapter includes ten sections that each discuss an agreement that is created based on a scenario of applying an FDT. These scenarios are created based on the requirements discussed in Chapter 3. Chapter 5 discusses a potential data access matching algorithm. Chapter 6 presents the results of the treatment validation phase. Chapter 7 gives a conclusion that answers the main research question. Moreover, a discussion, limitations of the research and topics for future work are discussed in the chapter. The discussion is about whether automation is necessarily a good choice for data access. The research aims to contribute to the field of data access automation and FDTs.

2. Literature review

This chapter analyses some of the relevant literature and related work that have been selected for this research. The chapter discusses comparable RELs, comparable data access automation tools and recent developments on ODRL related research.

2.1 Comparable RELs

As mentioned before, ODRL is the selected REL for this thesis due to a number of advantages it has, like its semantic richness, extensibility, FAIR compliance, W3C recommendation status, and the ability to create access conditions as well as a request. However, there are many more RELs that could have been chosen, some of the most prominent ones are described in [10] and [11]. The four most prominent RELs according to these papers are: Creative Commons, METSRights, MPEG-21 and XACML. This section describes these RELs and their benefits and drawbacks.

(1) Creative Commons
Initially developed in 2002, Creative Commons (CC) provides a framework for articulating the rights associated with open access web resources, including HTML documents, RSS feeds, and digital audio files [12]. The CC licence is machine-readable, as it exists in the form of a digital document. However, it lacks machine-actionable control over the usage of licensed content; CC depends on a system of trust and existing copyright laws to safeguard digital content. When a CC licence is assigned to a resource, a CC graphic displayed on the web page or embedded within the resource (for instance, in digital audio files) links to the Creative Commons website, which hosts the rights expression. The CC licence itself is a concise version of more detailed licences available on the CC website, which includes comprehensive legal terminology for the few licence statements provided. The Creative Commons licensing framework is loosely based on the open licensing model of the Free Software Foundation's GNU General Public License.

The CC metadata record comprises two components: the work and the licence. The work section utilises simple Dublin Core metadata elements to describe the item to which the licence applies. Given that the CC licence is designed to be universally accessible to all Internet users, the user documentation for the CC licence is notably straightforward. A single web page offers a "fill in the blank" function, enabling anyone to create a licence easily.

(2) METSRights
METSRights describes itself as a simple rights schema that could be used while the more comprehensive RELs are being developed and debated [13]. The REL focuses on 3 things [13]:

- Digital resources owned or controlled by a digital repository rather than e-resources accessed remotely, formally licensed and subscribed to by an organisation (the area covered by the DLF ERMI group).
- Declaring the rights holders and rights associated with the digital resources mentioned above rather than trying to fully express all rights as would a REL designed to be used with a Digital Rights Management system or product.

- Simplifying the declaration as much as possible given that the whole DRM & REL scene is changing so rapidly.

The schema itself is based on XML and has 3 main elements: (1) simple declaration of the type of right and the public statement of that rights declaration, (2) the naming of the rights holder with appropriate contact information, and (3) the context for the rights declaration based on the type of user who has a set of permissions for a digital object.

(3) MPEG-21

MPEG-21 is a comprehensive suite of standards concerning digital multimedia resources, encompassing seven primary components: identification of digital items, content representation, delivery protocols, and intellectual property management [14]. The intellectual property management aspect includes a REL and a framework for developing a data dictionary. The entirety of the MPEG-21 standard has been adopted as ISO 21000. Part 5 of ISO 21000/MPEG-21 details the REL, while Part 6 outlines the structure for a data dictionary to support the REL. Although the standard contains entries for key verbs used in the REL, a complete data dictionary for Part 5 has not been fully developed.

The MPEG-21 REL is specifically designed for licensing digital materials, with a focus on video and audio (MPEG stands for "Motion Picture Experts Group"). The REL was developed by the MPEG-21 standards group based on the Extensible Rights Markup Language (XrML). ContentGuard, the company owning and managing XrML technology, participated in the development of the MPEG-21 REL. This standard is intended to be unambiguously machine-actionable, enabling interaction with software and hardware that enforce the licence permissions. It anticipates future implementation of trusted systems technology, which will allow end-to-end control over digital works from publication through distribution to the end user's device.

Although the MPEG-21 standard was created mainly by representatives of the multimedia intellectual property industries, the REL standard was intentionally designed to be broad, making it applicable to a wide range of digital products. Consequently, the standard is highly generalised, requiring that actual implementations use MPEG-21 Part 5 as a meta-language to create specific functions needed for various applications. For instance, an extension to MPEG-21 Part 5 for e-books is currently under consideration by the Open eBook Forum, an industry group that develops standards for e-books. Should this standard be adopted, MPEG-21 Part 5 could become the primary REL for the trade e-book publishing industry in the United States and potentially in Europe.

(4)XACML

According to [11], XACML intends to manage secure authorization and to connect authorised users to resources through formalised policy statements. It is XML-based and its web site [15] provides a detailed description of the intended use of the framework and its rules. XaCML has its logic described in [16]. Even though it is not a REL [11], it is still included in the literature, since it is a system to manage access to data.

The basic model of XACML can be found in Figure 5. The Policy Enforcement Point (PEP) represents for example a filesystem or web server. The subject requests access to the PEP. Then the context handler creates a request based on the original access request from the subject. This request is then sent to the Policy Decision Point (PDP), which extracts the relevant policies from the Policy Administration Point (PAP) and it evaluates if the access request complies with the XACML policies. If the access request is in line with the policies then the subject gets access to the demanded data.

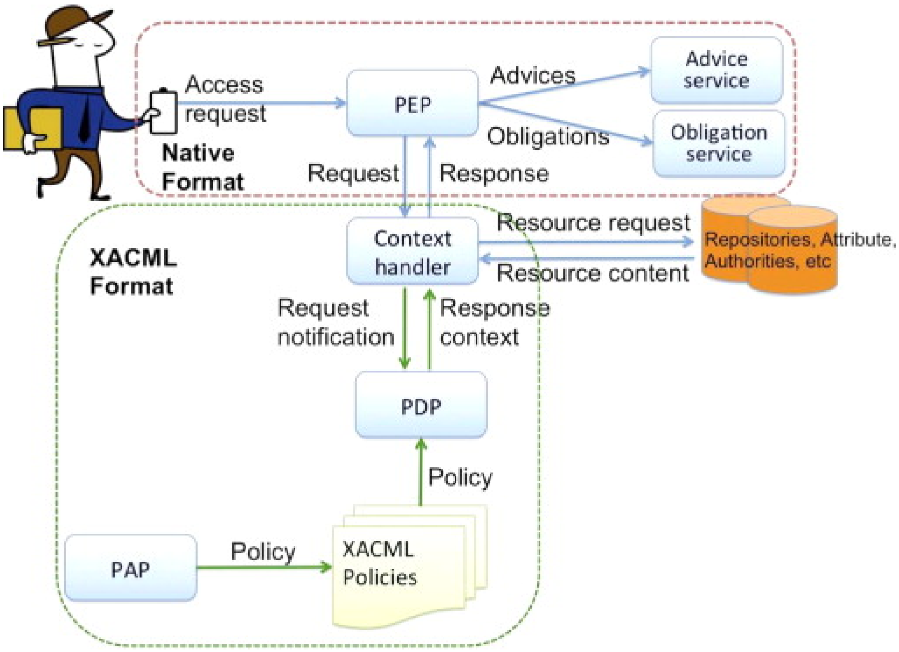


Figure 5: XACML basic model [16].

(5) REL choice

Even though all of the RELs have their own benefits and preferred use cases for which they were developed, it appears that ODRL is the most suitable for the context of a FDT, because of its semantic richness. It appears to be the only REL that is semantically rich, and complies with the FAIR principles. Furthermore, it is the W3C recommended REL. Therefore, it is a commonly used standard and likely interoperable with many other systems. Moreover, the REL should be able to provide the users with the ability to create access conditions as well as an access request. Most of the aforementioned RELs fail in this regard, except for ODRL and XACML. Although it is worth noting that even though XACML is technically not a REL, it is also suitable in the context of a FDT. However, it is not semantically rich enough to meet the FAIR principles and therefore it was not chosen.

2.2 Access automation Tools

The field of access automation tools is novel. However, some research has already been conducted on this topic. The most relevant access automation tools that were developed in recent years are discussed in this section, which are DALICC and DUO.

DALICC

The Data License Clearance Center (DALICC) is a software framework designed to facilitate the automated clearance of rights, thereby enabling the legally secure and time-efficient reuse of third-party data sources [17] [18]. The framework provides various Application Programming Interfaces (APIs) that grant access to a comprehensive licence database along with additional functionalities to streamline the licensing process for derivative works.

DALICC uses parts of ODRL, extends this in combination with other vocabularies, and also includes their own created semantics. The main functionalities of the DALICC framework are displayed in Figure 6. These functionalities include: the option to create a customised licence, the ability to detect licence conflicts, the ability to attach licences to an asset, and the ability to select a standard licence from a licence library. However, the downside of DALICC is that the creation of customised licences does not offer a lot of options and is therefore not applicable to specific use cases. However, DALICC is a suitable framework to use for generic licences.

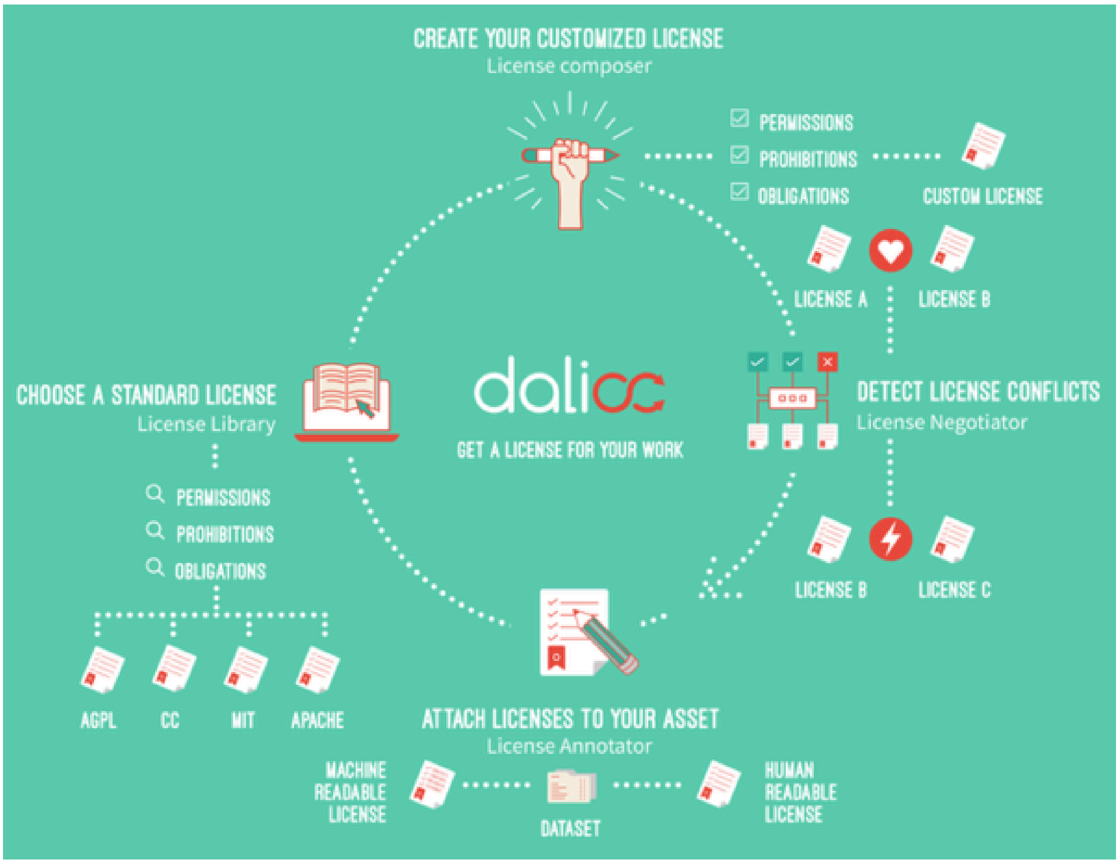


Figure 6: The DALICC framework and its functionalities [18].

DUO

Patient and participant consent forms frequently employ many different terms to describe the permissible uses and reuses of generated data. This lack of commonly agreed terms and definitions complicates the task for data access committees (DACs) to efficiently and confidently grant researchers access to data. In response, the Data Use Ontology (DUO), developed by the Global Alliance for Genomics and Health (GA4GH) Data Use & Researcher Identities (DURI) Work Stream, offers a standardised set of terms for tagging datasets with ‘use’ permissions [19] [20]. This ontology aids researchers in data discovery and streamlines DAC decision-making in the data access process. Figure 7 shows the process that DUO follows. The process includes 4 phases, which are the discovery phase, request phase, matching phase and access phase. The request can be created with DUO terms, which mostly cover health topics and are not generic enough to be able to be applied to a FDT in other domains.

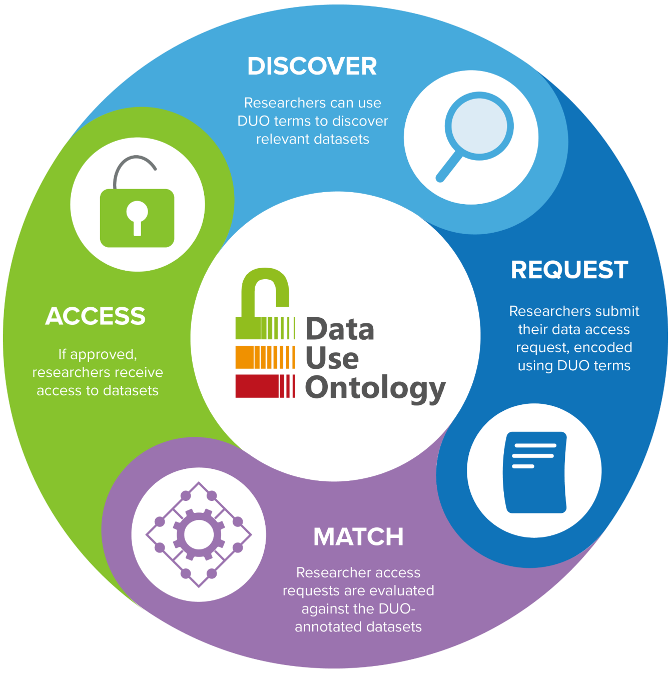

DISCOVER
Researchers can use DUO terms to discover relevant datasets
REQUEST
Researchers submit their data access request, encoded using DUO terms
MATCH
Researcher access requests are evaluated against the DUO-annotated datasets
ACCESS
If approved, researchers receive access to datasets
Data Use Ontology

Figure 7: The DUO process [20].

2.3 ODRL Advancements

Currently advancements are being made on ODRL. Researchers from the University of Madrid are currently working on an ODRL Translator and policy creation algorithm [30]. The ODRL translator is able to translate ODRL policies created in JSON-LD to human readable text. This will be useful for people who are not familiar with ODRL and want to make use of the technology. The algorithm is still undergoing development and currently is not usable yet. There are plans to incorporate AI technology in the future to enhance the accuracy of the translations.

Furthermore, the ODRL translator offers the option to define policies. First a policy description must be provided, which includes the type of ODRL policy (e.g. a request, offer etc.), the creator of the policy, a creation date and a description of the policy. Then rules can be added to the policy based on the Common Conditions of Use Elements (CCEs) [29]. The CCEs are the most popular elements in data access conditions in the healthcare sector. Additionally a rule type should be added, after which specific values about that rule can be added. After that, many more details can be added based on the needs of the user. Finally, once all the details have been entered, a JSON-LD ODRL policy can be displayed that includes everything that the user has added.

3. Design Cycle

The aforementioned Design Cycle [21] is elaborated upon in this chapter. The chapter discusses the problem investigation phase, the treatment design phase and the treatment validation phase and their relevance for the thesis. Figure 8 shows the applied methodology, which is based on the Design Cycle by Wieringa [21]. It starts off at the problem investigation phase in which the main problem is identified. Then an artefact is designed to solve the problem in the treatment design phase. Finally, the artefact undergoes validation during the treatment validation phase to determine whether it effectively addresses the identified problem. Should the artefact prove invalid, necessary modifications must be implemented. Consequently, the process reverts to the treatment design phase, followed by subsequent revalidation of the revised artefact. This iterative process continues until the artefact is confirmed as valid.

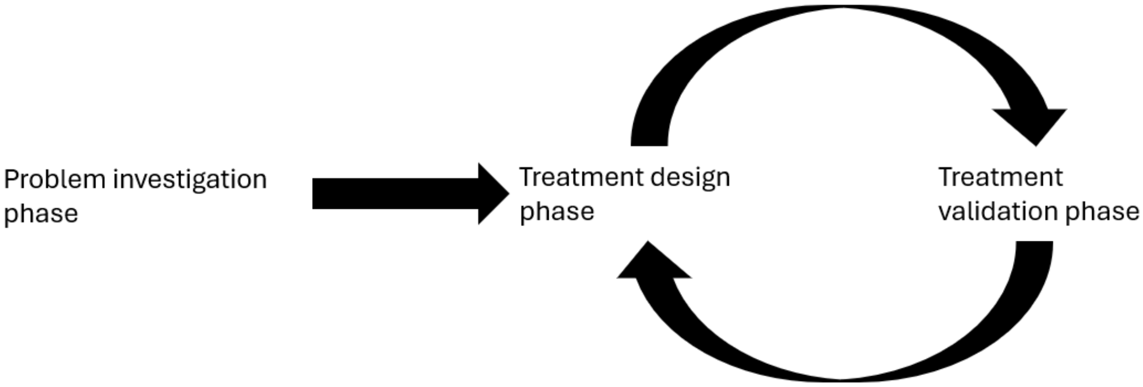


Figure 8: The applied methodology.

3.1 Problem investigation

In this phase of the design cycle, the main stakeholders are identified with their respective goals. Furthermore, the phenomena that caused the problem are identified. The identified problem is the currently inefficient process of data access. Furthermore, this is usually the phase in which knowledge questions are defined, but these are already defined in Section 1.1.

In the case of the FDT, anyone interested in some data can be a potential stakeholder. Anybody can launch a train or own data, whether they are an individual or an organisation. However to narrow it down, we define two main stakeholder groups. The main stakeholders are data owners and researchers. Data owners are the owners of an asset and are the party in charge of creating a data access policy. Researchers are the party that attempts to access the data for research purposes and create an access request.

The goals of the stakeholders mostly differ, but they have one in common, namely that both of them have interest in gaining knowledge or insights on their data. Researchers can use this knowledge for their research and data owners could use it for business related insights. Furthermore, data owners are interested in the security of their data. They have to comply with various laws, like the GDPR in Europe, for example. Therefore, they do not want their data to be misused, since they could get into legal trouble. Furthermore, they want to keep ownership over the data even after sharing the data, since it might take a lot of effort and resources to collect data for data owners. Hence they do not want everyone to claim ownership over their data.

However, the researcher has different goals in mind. They want to extract results from the data without modifying the data on the data owner's end, using an FDT. Furthermore, the researcher wants to access data for their research in compliance with the access conditions. In this way, they do not violate the access policy from the data owner and do not get into legal trouble, for example. The researcher wants this process to be as simple as possible, without having to read several documents, which may slow down data access.

Figure 9 shows an Archimate model of the stakeholders and their respective goals. The FDT with ODRL matching algorithm is a solution that fulfils these goals of the researchers wanting to access data and the data owner wanting to share their data.

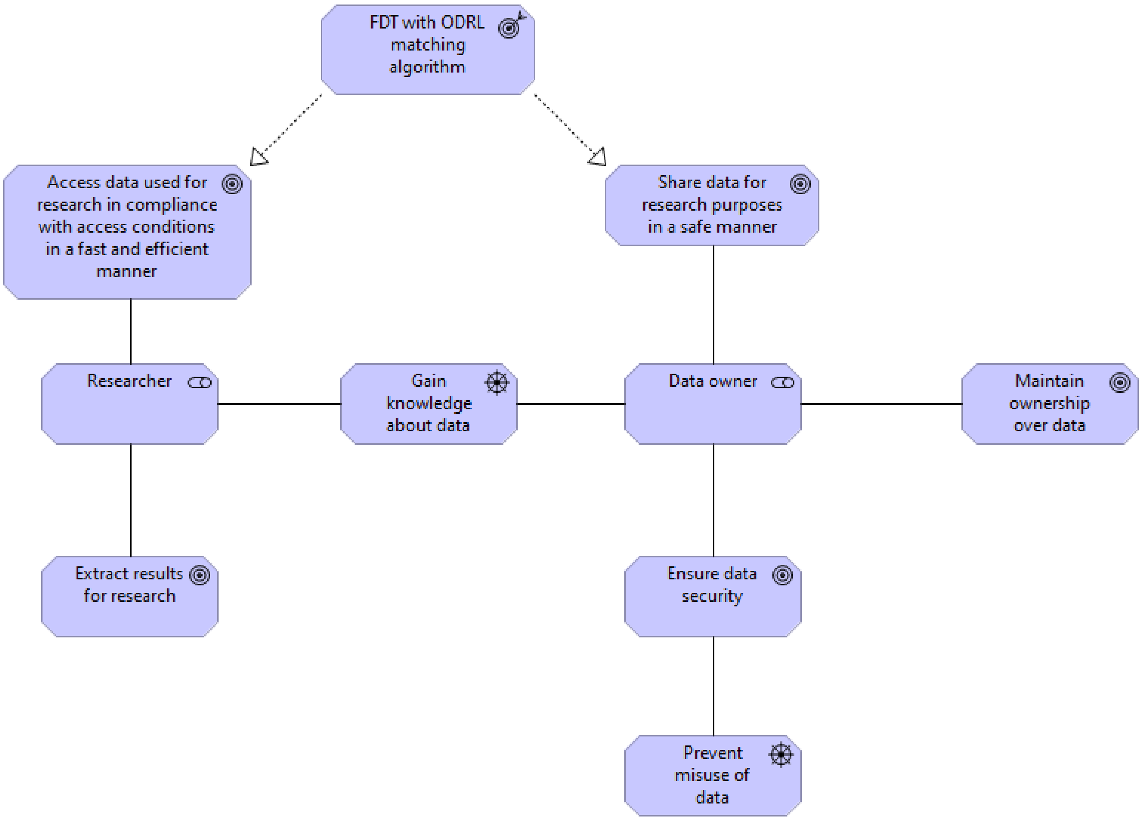

FDT with ODRL matching algorithm
Access data used for research in compliance with access conditions in a fast and efficient manner
Share data for research purposes in a safe manner
Researcher
Gain knowledge about data
Data owner
Maintain ownership over data
Extract results for research
Ensure data security
Prevent misuse of data

Figure 9: Archimate Goal model of the stakeholders involved.

3.2 Treatment design

In this phase of the Design Science Cycle, requirements are specified for the artefacts. The first artefact is a set of scenarios for ODRL agreements. Scenarios have to be designed to validate what the outcome of the algorithm should be based on a data access request and data access conditions. Furthermore, they should highlight the most common and useful ODRL aspects. The second artefact is an algorithm that has the ability to generate an ODRL agreement and grant access to an asset. Furthermore, the algorithm should also be able to deny access to an asset and notify the researcher why their access was denied. Below, the specific requirements for scenarios for ODRL agreements and a matching algorithm are defined.

Requirements for scenarios for ODRL agreements
The various scenarios for ODRL agreements have to be based on common real life scenarios of applying an FDT. The scenarios have to become increasingly complex to showcase the various aspects of the ODRL information model and vocabulary [16] [22]. Each scenario should highlight one aspect in particular that is not mentioned in the other scenarios. Furthermore, the scenarios should be easy to understand for researchers. The scenarios should incorporate an ODRL access policy, an ODRL access request and if these two match an ODRL agreement all denoted in RDF TTL format. Some more complex scenarios should also showcase the extensibility of ODRL by incorporating other ontologies. We assumed that access policies are always created before a request can be made, and that a request has to be modified to fit the access conditions to access data, and not vice versa. Lastly, the scenarios should form a basis for a standardised machine-readable access request that could be modified to fit the needs of a researcher.

Requirements for a matching algorithm
The algorithm should be able to compare an ODRL data access request with an ODRL data access policy. The comparison should then have one of two outcomes: either the request and conditions have a match or a mismatch. In the case of a match, it means that all of the rules and constraints of the data access request match the rules and conditions of the access policy. Furthermore, the access request and the access policy should be about the same target asset. In this case, an ODRL agreement between both parties should be automatically generated and the FDT gets automated access to the data and can run its algorithm. In the case of a mismatch, no ODRL agreement should be created and the FDT cannot access the data to run its algorithm. Furthermore, a message should be returned explaining why a mismatch occurred, so that the person that deploys the FDT can re-evaluate their access request and potentially change it to suit the access policy. An example could be that the data access conditions specify a date until the data can be used, but this is not specified in the data access request. Therefore this should be added to the data access request and should comply with the value in the access conditions, so that access can be granted. Furthermore, the conditions and access request should both be denoted in TTL RDF format, since this is the format expected by the algorithm.

3.3 Treatment validation

The treatment validation phase is an essential phase of the Design Science Cycle. In this phase, the artefact that was designed undergoes a validation process to confirm that it fits the specified requirements from the treatment design. If the artefact is not deemed suitable, the artefact should be re-designed based on the feedback from this phase. An artefact can be validated in many ways, for example, through an expert interview or an open case study [21].

In the case of this research, the RDFs representing the scenarios have been validated and also the scenarios. The created RDFs should be validated both on their structure as well as their semantics. The structure of the RDF can be validated using an RDF validator, which will check if the structure is correct, so that if certain characters are expected and not present, an error message is accordingly given. The semantics are a bit more complex to validate. When the ODRL translator [30] described in Section 2.3 is fully developed it could be used to translate the RDF back into natural language, and the natural language and the scenario of the use case can then be compared for correspondence. Furthermore, an ODRL expert could take a look at the RDFs and give their opinion on it. However, the ODRL expert might not be familiar with other vocabularies within the ODRL. Moreover, an AI tool can be asked to interpret the RDF. Since the RDFs are meant to be machine-readable, an AI tool should be able to interpret the offers and requests and should be able to decide whether an agreement should be created. The latter is our chosen validation method and ChatGPT 4 is used to validate our RDFs [31].

Furthermore, the scenarios themselves have also been validated. This has been done by conducting a survey on stakeholders in a real life FDT use case that the scenario from Section 4.8 is based upon. Since the stakeholders have experience with the FDT and offers and requests, they were able to tell whether the scenarios are relevant and realistic. Furthermore, these stakeholders also reviewed the created RDFs and gave feedback on them. In case that it appears that a designed artefact is not suitable according to the validation, the artefact should be redesigned and revalidated. However, this was not the case in this thesis.

4. Scenarios

In this chapter the results of applying the Design Science Cycle are discussed. In total ten scenarios were designed to fulfil the requirements that were defined in Chapter 3. The scenarios lead to ODRL agreements based on an offer and request. In this chapter we explain the RDFs that were created based on the scenarios. The most important elements of these RDFs will be explained and how to apply them. Furthermore, we will explain how an ODRL agreement can be constructed step-by-step.

As mentioned before in Section 3.2, scenarios were defined to identify the most common and relevant aspects of the ODRL information model and vocabulary [16] [22]. Furthermore, these scenarios were created to show how to apply ODRL and how it could possibly be extended with other ontologies. Moreover, they serve to show what an ODRL agreement should look like based on the access policy and access request, which would be the desired outcome of a matching algorithm based on the access policy and access request. In the RDFs created for the scenarios, the Unique Identifiers (UIDs) that have been created are based on “http://www.example.org/”, which means that they are not real UIDs. In a real life scenario, actual UIDs have to be created that replace these fictitious UIDs. Lastly, all of the scenarios have been validated with an RDF validator [23] checking their structure, but not their semantics. The created RDF documents can be found in Appendix A, except for scenario 4.1, which is included in the text. Furthermore, Appendix A includes images displaying how the data is linked in the RDF documents.

4.1 An open policy

The first scenario consists of an open policy. It is the most simple scenario, since there are no constraints or rules placed in the access policy. This means that anyone is free to use the data of the data owner if they send a request. The RDF representation of the scenario can be found below this paragraph. The RDF representation shows the most basic form of agreement, where there is a policy and a request and they create an agreement, because the policy and request are matching.

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .

ex:offer a odrl:Offer ;
        odrl:uid <http://example.com/offer:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:action odrl:use ;
        ] .
```

```
ex:request a odrl:Request ;
        odrl:uid <http://example.com/request:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:use ;
        ] .

ex:agreement a odrl:Agreement ;
        dcterms:references ex:offer, ex:request ;
        odrl:uid <http://example.com/agreement:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:use ;
        ] .
```

Every ODRL policy should always mention an asset, which should preferably be denoted with UID. Furthermore, every ODRL policy must have a UID, which identifies the policy and makes it possible to refer to it. Lastly, every party should have a UID that identifies the party.

The assigner is the party that creates an offer to share their data with assignees, so this is the data owner. An offer must always contain at least one odrl:permission or odrl:prohibition and a party with an odrl:assigner function.

The assignee is the party that creates an access request to match the offer of assigners, so this is a researcher. A request must always contain a target asset, a party with an assignee function and at least one permission or prohibition rule. A request can also contain an assigner if this party is known. However, to keep the RDF representation as simple as possible this was not included in this scenario. Furthermore, the RDF representation contains prefixes that are not used, these will however be used later in other scenarios and are included for that reason.

Since the offer and request match based on the fact that they have the same permissions and refer to the same asset, an agreement is made. An agreement always uses dcterms:references and refers to the offer and request using this. The agreement itself must always have a UID. The agreement contains a permission that states the assigner, assignee, target asset and the action that was agreed upon. In this case the most generic action is applied, which is ‘use’. Use can refer to many more specific actions, for example: printing the asset, sharing the asset with third parties, installing the asset, etc. Figure 10 shows a picture of the agreement created in this scenario and how the triples in the RDF look like. The figure was created using isSemantic [24].

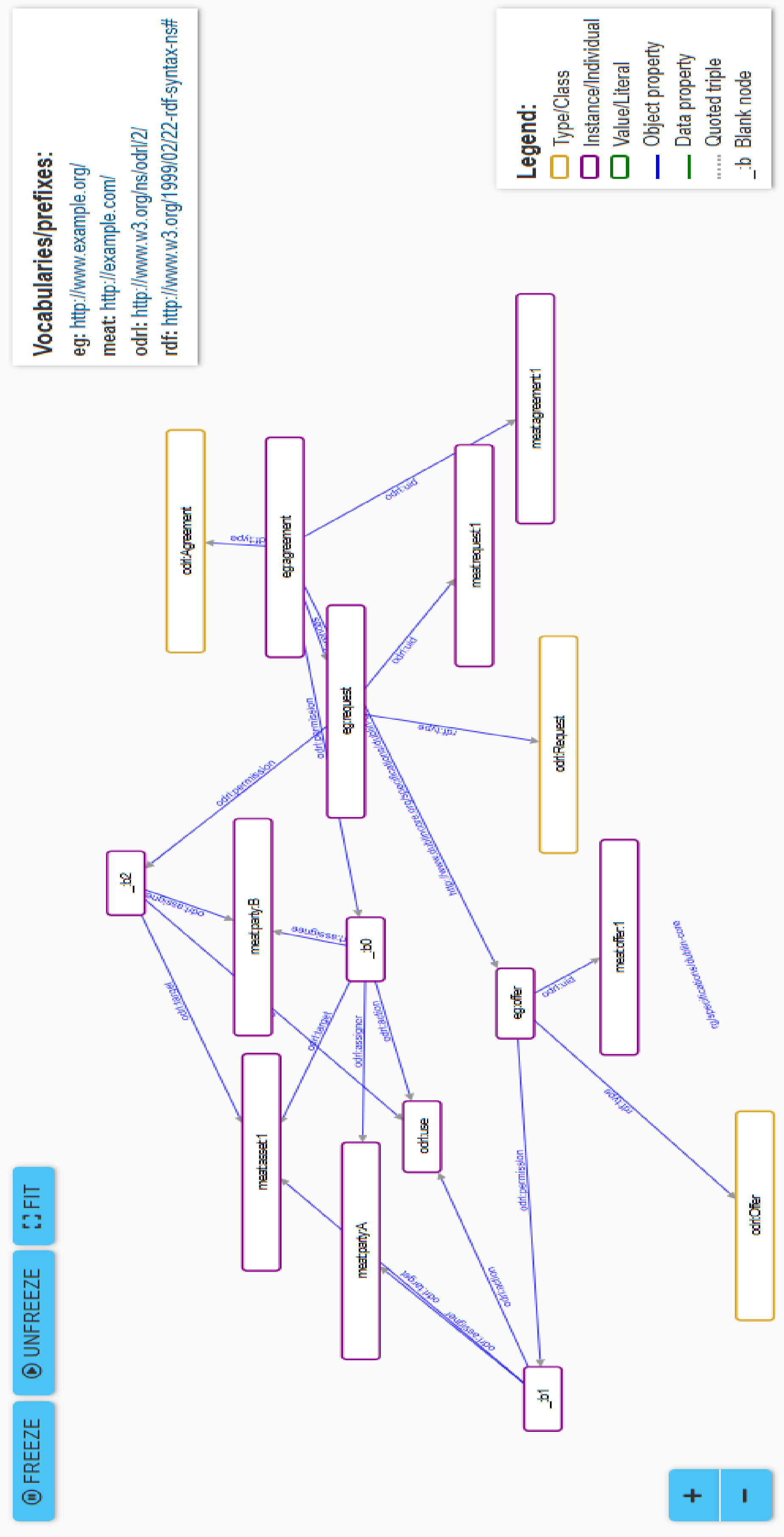


Figure 10: The linked data behind an ODRL agreement based on an open policy.

4.2 Compensation involved

This next scenario involves a data owner that is willing to share their data against monetary compensation in Euros. It is plausible that businesses that have spent resources to collect data do not want to give away this data for free. Instead they could be looking for compensation in return, possibly in the form of money. In this scenario, the data owner wants to share their data against a compensation of 100 Euros that has to be paid by the researcher before access is granted to the researcher, but this number can be replaced by any desired amount.

In essence the RDF documentation is similar to the previous scenario. However, now a duty has been added that contains an action with a constraint. Including a duty in a policy means that there is an obligation to perform the action that is contained in the duty by the assignee. Inside this duty there is the action, which instead of ‘use’ is now changed to ‘compensate’. The action is followed by a refinement.

A refinement is a method of further specifying the semantics of an action. In this case the refinement acts similarly to a constraint, because it contains a ‘leftOperand’, ‘operator’ and ‘rightOperand’. Instances of the ‘leftOperand’ class are used to define the expression in the constraint. The ‘operator’ is used to represent relational operators between the ‘leftOperand’ and ‘rightOperand’. The ‘rightOperand’ is the value that the ‘leftOperand’ has to be compared with using the ‘operator’. For further clarification, the ‘leftOperand’ in this refinement is ‘payAmount’, which defines an amount that should be paid. Moreover, the ‘operator’ is ‘eq’, which means equals. Lastly, the ‘rightOperand’ is ‘100,00’, which is a decimal value as specified behind. So this means that the amount that should be paid equals 100,00. After that, the unit of currency is defined and there is a link to a webpage that describes the concept of Euro. So the amount that has to be paid to get access to data equals 100 Euros based on this refinement.

Furthermore, a constraint has been added within the same duty. Constraints are a boolean expression that refine the semantics of an action or declare the conditions applicable to a rule. Constraints, as mentioned before, also contain the ‘leftOperand’, ‘operator’ and ‘rightOperand’ specifications. In this case, this constraint was added to specify that the duty must be completed before access can be granted to the asset. This means that the assignee must pay 100 Euros upfront to get access to the data. This is done by including the following piece of code into the duty:

```
odrl:constraint [
                    odrl:leftOperand odrl:event ;
                    odrl:operator odrl:lt ;
                    odrl:rightOperand odrl:policyUsage ;
             ] ;
```

The event which is referred to in this case is the policyUsage, which means the use of the policy. The ‘operator’ is set to less than. Therefore, this means that there is a constraint on the action that the event must be less than the policy usage, which means that the action has to be performed before the policy applies. This constraint will be added more times in later scenarios and is essential in some ODRL policies.

Because the offer in the RDF specification matches the request, an agreement is made that states that the assignee has to pay the assigner 100 Euros before they can access the data.

4.3 How to apply a role

This scenario is about a data owner who demands that only students are able to gain access to their data. This means that the data owner does not want businesses or regular citizens to be able to access their data. This involves the use of another vocabulary besides ODRL, namely ‘vCard’. The vCard ontology is meant for describing people and organisations and is the recommended way of describing more details about a party by ODRL.

In this scenario, the ODRL offer should specify that the assignee should have a role of ‘student’. This is done by specifying an assignee, with a ‘vCard:Role’ that refers to the Wikidata page that describes what a student is. The request should then also state that the assignee is a student by linking to the same Wikidata concept. When these steps are applied and both the offer and request match, an agreement is made that includes that the assignee is a student. This gives permission to the researcher to use the data. In other scenarios, it could also be specified that the assignee must be a doctor, for example. This can be done by changing the Wikidata link that follows the ‘vCard:Role’

4.4 The use of a specific action and prohibition

This scenario is about the use of a specific action and a prohibition. The scenario is about a data owner that only allows a researcher to copy their data and does not allow modification of the data, since the data owner wants the data to remain authentic. This means that the generic action ‘use’ can not be applied in the offer, because then the researcher would be able to do anything with the asset, like, for example, deleting it. However, ‘use’ has to be replaced with the specific action ‘reproduce’, which means that the assignee is allowed to create duplicate copies of the asset. This is included in the permission.

Furthermore, the assignee is not allowed to modify the data in any sort of way. The way that this is achieved in ODRL is by adding a prohibition. By including a prohibition in an offer, the data owner restricts the researcher by giving them the inability to perform certain actions over an asset. In this case, the action ‘modify’ is included in the prohibition, which hinders the researcher from modifying the data. Since both the offer and request agree on this, an agreement is made that includes both the permission and prohibition.

4.5 Basic constraints

This scenario includes the use of basic constraints. It is about a data owner that only allows their data to be used within the Netherlands up until the 31st of December 2025. Therefore the offer and request should contain a duty with time and spatial constraints, to ensure that the researcher is aware of these conditions.

As mentioned before, a constraint contains a ‘leftOperand’, ‘operator’ and ‘rightOperand’. The constraints should be included inside a ‘Duty’ in the access policy, since the assignee is only permitted to use the asset when they agree to the obligation of applying the constraints. The following code snippet from the RDF shows how the ‘Duty’ is supposed to look like:

```
odrl:Duty [
        odrl:constraint [
                odrl:leftOperand odrl:spatial ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "NLD" ;
        ] ,
        [
                odrl:leftOperand odrl:dateTime ;
                odrl:operator odrl:lt ;
                odrl:rightOperand "2026-01-01"^^xsd:date ;
        ] ;
```

The first constraint is a spatial constraint and it states that the spatial location should be equal to “NLD”, which is the ISO3166 country code [25] for the Netherlands. The use of ‘odrl:spatial’ recommends values in the ‘rightOperand’ to be equal to an ISO3166 country code. In this scenario, the country code of the Netherlands is chosen, however this can be changed to any other country/continent/region accordingly.

The second constraint is a dateTime constraint, which ensures that the assignee is allowed to use the data up until the date or time. In this scenario, the data can be used up until the 31st of December 2025, since the dateTime should be less than the 1st of January 2026. When a specific time should be added, Timezone information should be included, however, this is already constricted in this use case by limiting the ability to use the data in the Netherlands.

4.6 Example of a mismatch

This scenario is about a mismatch and is based on the previous scenario. Again a data owner wants the researcher to use the data before the 1st of January 2026 within the Netherlands. However, this time the researcher has made a request stating that they want to use the data in Brazil up until the 31st of December 2026. In this case, the offer and request do not match and thus no agreement is created.

In the case of a conflict between the offer and the request, the default strategy specified by ODRL is always to invalidate the agreement. However, this can be overridden by specifying a conflict strategy in ODRL. For example, in case of a conflict between the request and offer, if the conflict is set to ‘perm’, so that access is still permitted in case of a conflict.

4.7 A combination of former scenarios

This scenario aims to combine some of the previous scenarios into a more complete and complex RDF. The scenario is as follows: a researcher wants to use some competition sensitive data from a business, the business wants to be compensated for this in the form $300, since it has cost them working hours to collect the data. The money should be paid before access to the data can be granted. The data owners only want researchers related to universities to be able to access their data, to protect their data from competitors. Furthermore, any personal data or data that is considered sensitive should be anonymized to comply with regulations like the GDPR.

As mentioned before, a compensation duty must be included in the agreement and this must be paid before, similarly to the scenario described in Section 4.2. Furthermore, a role is added to the assignee that specifies that the assignee should be a researcher. The Wikidata page for a researcher is used for this following the ‘vcard:Role’. Then a duty stating that a compensation of 300 US dollars is required before the policy can be applied. Then another duty follows that specifies that the data must be anonymized, by using the anonymization action. However, the anonymization does not have to happen beforehand, since this would be impossible. Since the offer and request match, an agreement is created.

4.8 A real life FDT use case

This scenario is based on a real life use case of an FDT, that is deployed to collect data from farmers about their crops to improve eating habits for regular consumers, specifically youngsters and people suffering from diseases. The idea is to link the farmers directly with consumers to remove the step of consumers needing to go to grocery stores to purchase products. Instead, the consumers can directly purchase necessary products from farmers. The system works as a recommender system, where consumers are recommended a recipe based on their eating habits and their preferences. The recipe serves as a basis for the products that are then recommended to the consumer. A farmer gets then recommended to the consumers based on certain factors, like the amount of energy and water they consume and how many nutrients there are in their products etc.

However, the farmers are reluctant to cooperate with the recommender system, since they are scared that the data that is collected will be used against them and that it damages their business. Furthermore, the farmers are scared that they might score worse than their competitors on the environmental key performance indicators for example. The farmers have indicated that they are willing to cooperate with the system as long as sharing their data will lead to an increase in profit. The problem with the farmers' requirements to cooperate is that it is hard to guarantee a direct increase in profit in an access request, but an increase in visibility leading to more customers and thus indirectly more profit can be guaranteed.

There are a few notable lines in the RDF specification of this scenario that are discussed. Firstly, the assigner is referred to as the wikidata page of a farmer, since we know that a farmer is the assigner in this case. Secondly, the data is used for commercial purposes as has been specified in the RDF specification. Furthermore, identity authentication is necessary before the asset can be used by the assignee. To give the assignee the duty to authenticate their identity, the ODRL action ‘include’ is used in combination with a refinement. The snippet below shows how identity authentication can be added as duty.

```
odrl:Duty [
        odrl:action odrl:Include ;
        odrl:refinement [
                odrl:leftOperand odrl:media;
                odrl:operator odrl:eq;
                odrl:rightOperand dpv:IdentityAuthentication ;
        ] ;
        odrl:constraint [
        odrl:leftOperand odrl:event ;
        odrl:operator odrl:lt ;
        odrl:rightOperand odrl:policyUsage ;
        ] ;
] ;
```

The refinement for the ‘include’ action means that a mediafile should be included as a separate asset that can be used for identity authentication. To refer to an identity authentication, the Data Privacy Vocabulary (DPV) [26] has been used. The Data Privacy Vocabulary is an ontology that expresses metadata about the use and processing based on legislative requirements.A piece of code used before is reused here, which is the constraint that specifies that the duty must be fulfilled before access can be granted to the access.

The DPV has been applied multiple times in this RDF. It has been used twice combined with the ODRL action. The action ‘SellProducts’ and also ‘marketing’ have been used to guarantee that the assigner has the duty to sell and promote the products of the farmers. In this way, the farmers might not be as reluctant to share the data, since it gives the system a duty to promote and sell the farmers products.

Lastly, some generic actions have been included that could aid in reducing the reluctance of farmers to share their data. These are the duty to inform the farmer of the results of the queries of the FDT and to share the source code of the query of the FDT. Even though the farmer might not understand the source code, it leads to an increased transparency if someone explains the source code to them, since they would be able to know specifically what happens to their data.

4.9 Non-specified asset

This scenario is a realistic one in which a researcher would like to apply an FDT. Since the researcher might not know what assets the data owner potentially has that contain the data they are interested in, instead of specifying a target asset, characteristics of the target asset are defined. This is done by applying the Shapes Constraint Language (SHACL) [27]. SHACL is used for defining the shape of an asset so that characteristics can be specified to which a target asset must possess.

In this scenario, a researcher wants to conduct research on specific personal health data and for their research they need personal health data that is related to blood. Therefore, they decide to run an FDT, trying to extract as much data related to blood from hospitals as possible. However, the researcher does not know what kind of assets the hospitals have that contain this data. To get access to the data, the hospitals require identity authentication from the assignee, which should be done beforehand. Moreover, the hospitals require the researcher to anonymize the data to comply with the GDPR. Lastly, the hospitals have added metadata to the access policy to describe the asset, to provide clarity to researchers.

The offer in this scenario talks about a specific asset, because the data owner or hospital in this case, knows what asset they have that contains the information that the researcher wants. In contrast, the researcher does not know what the specific target asset is, so they cannot refer to the UID of this target asset. Therefore, SHACL is applied to describe the target asset as shown in the following RDF specification snippet:

```
odrl:target [
        a sh:NodeShape ;
        sh:targetClass odrl:Asset ;
        sh:property [
                sh:path dcterms:conformsTo ;
                sh:hasValue wd:Q7873 ;
        ] ;
] ;
```

The target defined in the request of the FDT is an ODRL asset that has the properties of a path and value. The path of the asset is ‘dcterms:conformsTo’, which means that the asset should conform to a certain value. This value is then described in the line below, with the ‘hasValue’ specification. The value that this is referring to is the Wikidata link for blood. So in essence in this snippet of code, it states that the target should be an asset that conforms to anything related with blood. Figure 11 shows the SHACL structure supported by blank nodes, where b5 is the target node.

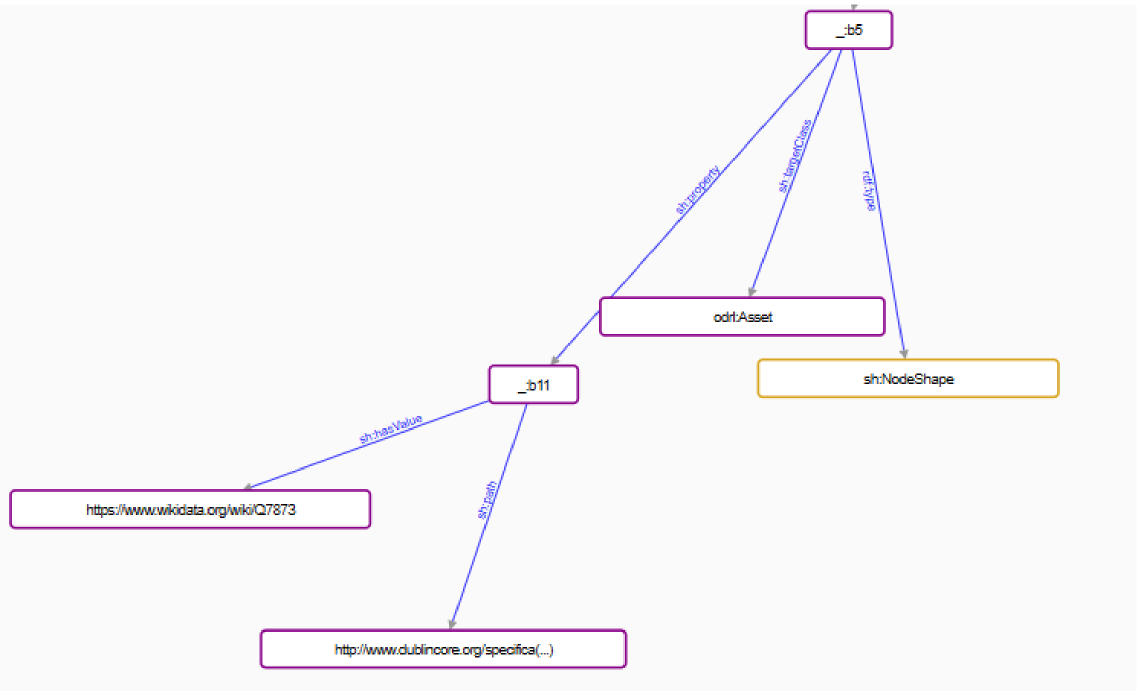


Figure 11: The structure of a SHACL asset.

The offer provides a description and an issue date in its metadata that refers to the target asset. This will help researchers in an FDT scenario, because when a researcher gathers an access policy, they will know what kind of data the asset contains. ODRL recommends applying the Dublin Core Metadata Terms (dcterms) [28] to describe the metadata for ODRL policies. The code snippet below shows how to implement the dcterms in ODRL:

```
dc:description "This rdf contains personal health data about people and their
respective blood types. It also has information containing if the person with the blood
type suffers from any diseases." ;
dc:issued "2020-01-01T12:00" ;
```

Only a description of the asset and the issue date of the policy are mentioned here, but many more things can be included. For example, the creator of the policy, the jurisdiction under which the policy is relevant, etc.

Because the target asset in the offer conforms to blood related data and the request targets such an asset as well, an agreement is created. In this agreement the SHACL code is not mentioned, since the target asset is indicated in the agreement. Because the request does not specify one target asset in particular, this request will likely lead to multiple agreements with different assets instead of only one. However, for the sake of the example, only one agreement is made.

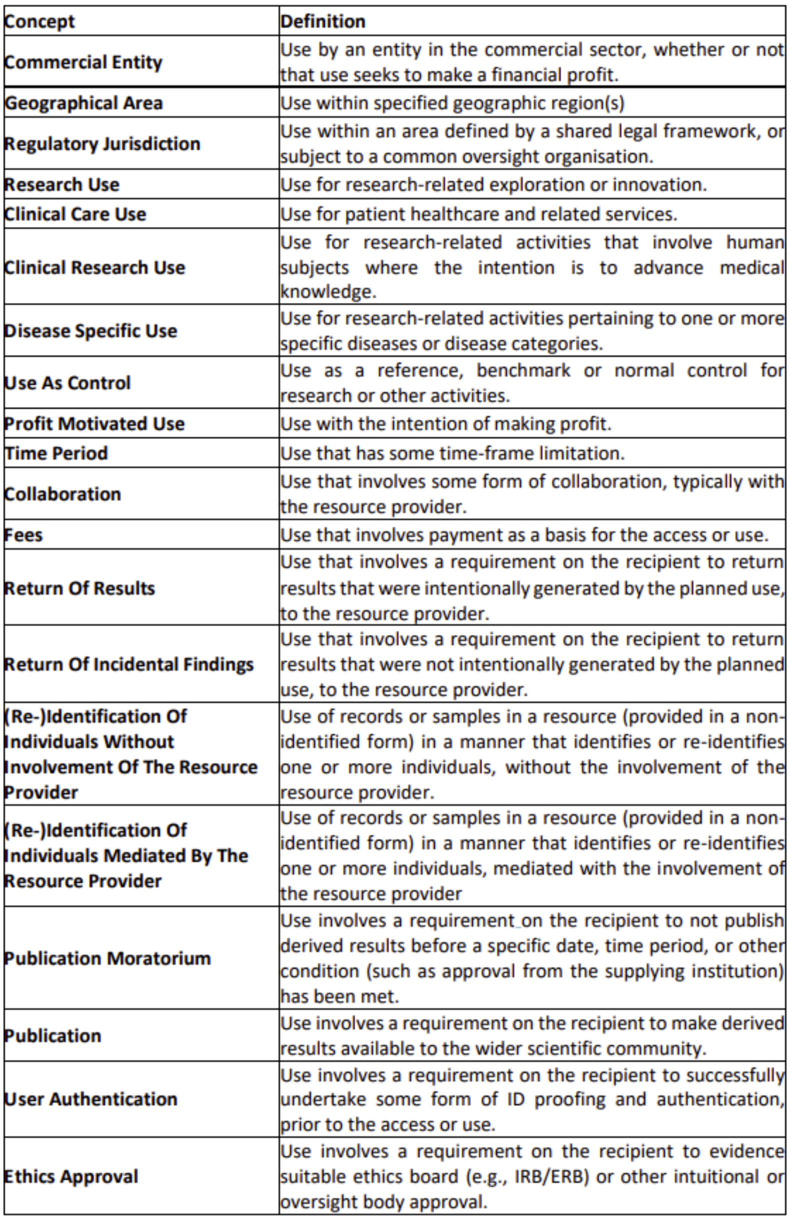
4.10 Standardised access request

Being able to produce standardised machine-readable access requests is one of the goals of this thesis. A standardised access request could then be modified by adding and/or removing certain elements based on the specific needs of the researcher or the access policy that was defined. However, this means that the main elements for most of the access requests are present. The access request of this scenario is based on the most common elements in health data access policies [29], since this data is often privacy sensitive and thus requires detailed access requests with many constraints. The most common elements in access policies of medical data can be found in Figure 12.

Table 1 shows which elements map to which part of the ODRL request. For values that are optional, placeholders are applied. For example, for the geographical area, the Netherlands was chosen in the RDF specification as a constraint. However, these placeholder values can be adapted accordingly. Regarding the potential research uses (clinical care use, disease specific use, etc.) they all relate to the same concept of using the data in a certain way, so this is reflected with the action:use in the RDF specification, use can be replaced with a more specific action based on the needs of the policy.

Some elements can be represented in different ways in the RDF specification. For example, the regulatory jurisdiction is represented in the same way as the geographical location. Because by restricting the geographical location, the regulatory jurisdiction is also restricted to a certain jurisdiction. However, if this is not the case, there is always room for a more refined specification by applying DPV, for example. Furthermore, the anonymization action is added even though it is not stated anywhere that this should be the case. The assumption is that the health data must be anonymized before it can be published, since it is personal data. Lastly, there is a duty for a researcher to share the source code with the data owner even though this is not specifically mentioned in the access conditions. However, to prevent accidental finding and to let a data owner know the goals of the researcher it would be helpful to include this.

When all of these elements are combined, an access request is designed to specifically match the access policy, after which an agreement can be attached. It is always possible to modify the access request based on the specific needs of a use case, but it should contain most of the required elements in access policies.



Concept	Definition
Commercial Entity	Use by an entity in the commercial sector, whether or not that use seeks to make a financial profit.
Geographical Area	Use within specified geographic region(s)
Regulatory Jurisdiction	Use within an area defined by a shared legal framework, or subject to a common oversight organisation.
Research Use	Use for research-related exploration or innovation.
Clinical Care Use	Use for patient healthcare and related services.
Clinical Research Use	Use for research-related activities that involve human subjects where the intention is to advance medical knowledge.
Disease Specific Use	Use for research-related activities pertaining to one or more specific diseases or disease categories.
Use As Control	Use as a reference, benchmark or normal control for research or other activities.
Profit Motivated Use	Use with the intention of making profit.
Time Period	Use that has some time-frame limitation.
Collaboration	Use that involves some form of collaboration, typically with the resource provider.
Fees	Use that involves payment as a basis for the access or use.
Return Of Results	Use that involves a requirement on the recipient to return results that were intentionally generated by the planned use, to the resource provider.
Return Of Incidental Findings	Use that involves a requirement on the recipient to return results that were not intentionally generated by the planned use, to the resource provider.
(Re-)Identification Of Individuals Without Involvement Of The Resource Provider	Use of records or samples in a resource (provided in a non-identified form) in a manner that identifies or re-identifies one or more individuals, without the involvement of the resource provider.
(Re-)Identification Of Individuals Mediated By The Resource Provider	Use of records or samples in a resource (provided in a non-identified form) in a manner that identifies or re-identifies one or more individuals, mediated with the involvement of the resource provider
Publication Moratorium	Use involves a requirement on the recipient to not publish derived results before a specific date, time period, or other condition (such as approval from the supplying institution) has been met.
Publication	Use involves a requirement on the recipient to make derived results available to the wider scientific community.
User Authentication	Use involves a requirement on the recipient to successfully undertake some form of ID proofing and authentication, prior to the access or use.
Ethics Approval	Use involves a requirement on the recipient to evidence suitable ethics board (e.g., IRB/ERB) or other intuitional or oversight body approval.

Figure 12: Most common elements in medical data access conditions and their descriptions [11].

Concept	ODRL
Commercial entity	odrl:constraint[odrl:leftOperand odrl:purpose ; odrl:operator odrl:eq ;

	odrl:rightOperand dpv:AcademicResearch ;]
Geographical area,Regulatory Jurisdiction	odrl:constraint [odrl:leftOperand odrl:spatial ; odrl:operator odrl:eq ; odrl:rightOperand "NLD" ;]
Research Use, Clinical Care Use, Clinical Research Use, Disease Specific Use, Use As Control, Profit Motivated Use, (Re-)Identification Of Individuals Without Involvement Of The Resource Provider, (Re-)Identification Of Individuals Mediated By The Resource Provider	odrl:action odrl:use
Time period	odrl:constraint [odrl:leftOperand odrl:dateTime ; odrl:operator odrl:lt ; odrl:rightOperand "2124-01-01"^^xsd:date ;]
Collaboration,Publication Mortuarium, Publication	odrl:Duty [odrl:action odrl:sharing; odrl:refinement [odrl:leftOperand odrl:dateTime; odrl:operator odrl:gt ; odrl:rightOperand "2024-05-01"^^xsd:date ;];] ; odrl:Duty [odrl:action odrl:anonymize ;] ;
Fees	odrl:Duty [odrl:action odrl:compensate ; odrl:refinement [odrl:leftOperand odrl:payAmount ; odrl:operator odrl:eq ; odrl:rightOperand "0,00"^^xsd:decimal ; odrl:Unit "http://dbpedia.org/resource/Euro" ;] ;];
Return of results, Return of incidental finding	odrl:Duty [odrl:action odrl:inform ;] ;

	odrl:Duty [odrl:action "http://creativecommons.org/ns#SourceCode" ;] ;
User Authentication	odrl:Duty [odrl:action odrl:Include ; odrl:refinement [odrl:leftOperand odrl:media; odrl:operator odrl:eq; odrl:rightOperand dpv:IdentityAuthentication ;] ;] ;
Ethics Approval	odrl:Duty [odrl:action odrl:obtainConsent;] ;

Table 1: Mapping the common elements of access conditions to ODRL.

4.11 Similarities of the agreements

From these scenarios and their respective RDF specifications we can conclude that there is a common method to create an agreement. Firstly, an agreement can only be made when a request matches the permissions and prohibitions of an offer. Secondly, dcterms:references is always used to connect the offer and request. Furthermore, the agreement should always copy the target asset based on the offer. Since a request does not have to specifically mention a target asset based on scenario 4.9, while an offer must always do this. Lastly, an agreement should always mention an assigner and assignee, which are the parties that are involved in the agreement.

5. Matching algorithm

This chapter describes how a matching algorithm should function based on the requirements specified in Section 3.2. The various steps that an algorithm should follow are explained in this section and a UML activity diagram is presented that visualises how the algorithm should work. Furthermore, this section compares the current situation with the desired situation of data access and how the desired situation improves upon the current situation.

Algorithm

Based on the requirements, a design for an algorithm is proposed. Figure 13 shows a UML activity diagram of the steps an algorithm should take to conform to the requirements. The activity diagram was created using Draw.io[2]. The algorithm should start as soon as the FDT arrives at a data station, then it should check if the target asset has access conditions in place. If this is not the case, the algorithm should already end, since the FDT should be able to access the asset without needing an access request.

However, if access conditions are in place, the algorithm has to check if the FDT has an access request included in its metadata. If this is not the case, the algorithm should end displaying a message that an access request should be included within the FDT, since there are access conditions present. It could even display the access conditions, to save time in the future when trying to match a request.

Then the algorithm checks if the request mentions the same target asset as the conditions. This check could be complex if the scenario is similar to the scenario of Section 4.9, since in that case, the algorithm should be able to compare SHACL nodes with the actual target asset. Because of this, the algorithm should be able to process SHACL in the access request. If there is a match between the request and conditions, the algorithm continues with the next step. On the other hand when there is no match, a message is displayed informing that the FDT should adjust the target asset or the characteristics of the asset.

Following those checks, the algorithm will need to check what types of permissions and prohibitions are present in the access request and access conditions. It needs to check the type of duties and constraints of these permissions and prohibitions. If these differ, then there should be no match automatically. The algorithm should then return a message specifying that the request should add/remove duties or constraints based on the types present in the access conditions. If there is a match between the types of duties/constraints, then the algorithm continues.

The next step is that the algorithm compares the values of these duties/constraints to see if they are matching. In case these are not matching like in the scenario of Section 4.6, a message should be returned stating the values of the access conditions and that the request should be adjusted to be granted access to the target.

[2] https://app.diagrams.net/

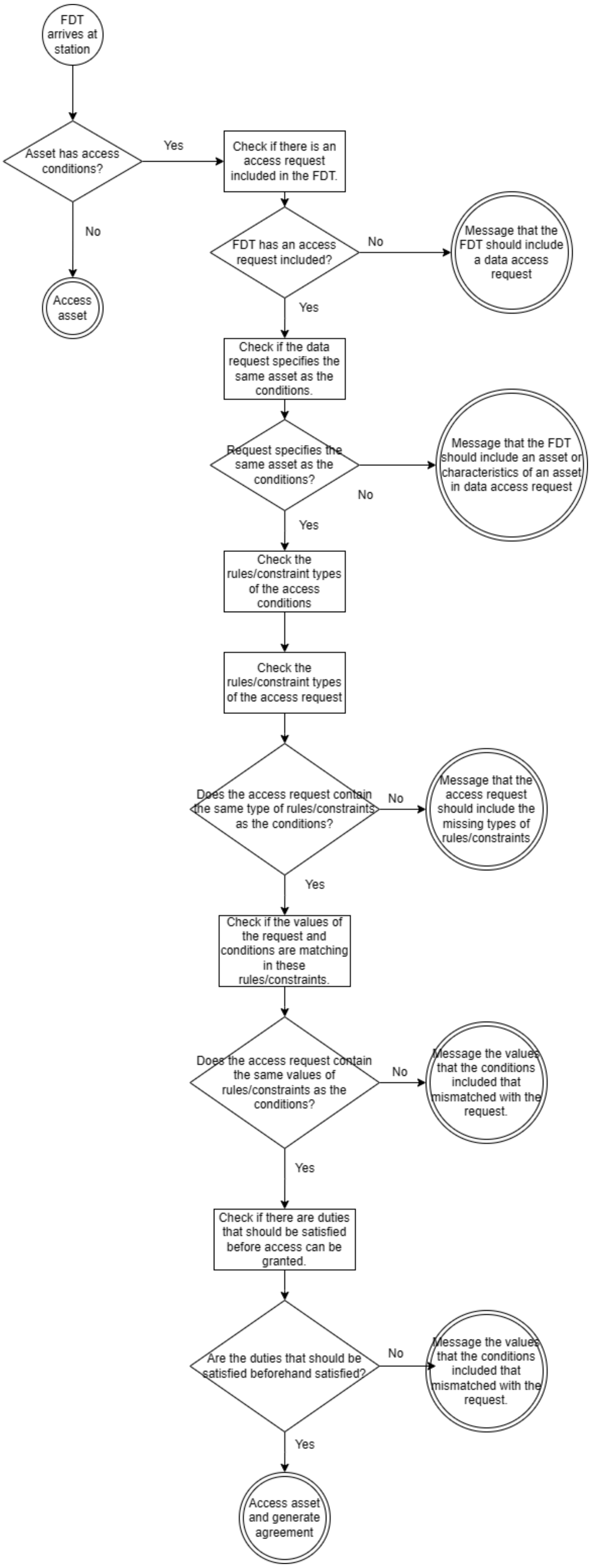

FDT arrives at station
Asset has access conditions?
Yes
No
Access asset
Check if there is an access request included in the FDT.
FDT has an access request included?
No
Message that the FDT should include a data access request
Yes
Check if the data request specifies the same asset as the conditions.
Request specifies the same asset as the conditions?
No
Message that the FDT should include an asset or characteristics of an asset in data access request
Yes
Check the rules/constraint types of the access conditions
Check the rules/constraint types of the access request
Does the access request contain the same type of rules/constraints as the conditions?
No
Message that the access request should include the missing types of rules/constraints
Yes
Check if the values of the request and conditions are matching in these rules/constraints.
Does the access request contain the same values of rules/constraints as the conditions?
No
Message the values that the conditions included that mismatched with the request.
Yes
Check if there are duties that should be satisfied before access can be granted.
Are the duties that should be satisfied beforehand satisfied?
No
Message the values that the conditions included that mismatched with the request.
Yes
Access asset and generate agreement

Figure 13: UML activity diagram of the matching algorithm

In case they match, one last check should be performed if a piece of code, like below, is included in the request and conditions:

```
odrl:constraint [
                    odrl:leftOperand odrl:event ;
                    odrl:operator odrl:lt ;
                    odrl:rightOperand odrl:policyUsage ;
             ] ;
```

In case this piece of code is included, an action should be performed before access can be granted. The algorithm should be able to check this, possibly by integrating some necessary third-party software. For example, this could be used for identity authentication, which is likely to be a common duty that should be completed before access can be granted for security reasons. If there are no duties that should be completed beforehand in place or these duties are satisfied, access should be granted to the asset and an agreement should be generated similarly to the agreements from Section 4.1.

Figure 14 displays a UML activity diagram of the desired situation of data access requests once the algorithm is fully implemented, which was created using Draw.io[3]. A data owner has data to share and creates an ODRL offer for an asset. Then a researcher should create an ODRL access request and add it to the metadata of an FDT. Following, the FDT is deployed and a data owner receives the FDT with the access request. Then the matching algorithm should run on the data owners side and the request and offer should be compared. If the data owner has no computational power for running the algorithm, the algorithm could even be dynamically staged in the cloud [5]. Based on the result of the matching algorithm, either access is granted or the researcher receives a message that access was denied, including the reason.

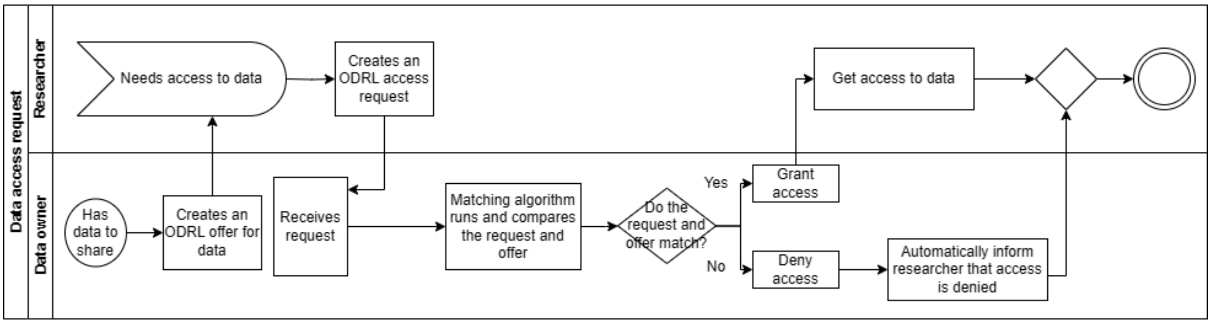

Data access request
Researcher
Data owner
Needs access to data
Creates an ODRL access request
Get access to data
Has data to share
Creates an ODRL offer for data
Receives request
Matching algorithm runs and compares the request and offer
Do the request and offer match?
Yes
No
Grant access
Deny access
Automatically inform researcher that access is denied

Figure 14: Desired situation of access requests after the FDT has been implemented.

Comparing Figure 14 with the current situation of Figure 3 in Section 1.1, the process should be significantly faster and more efficient. Less parties are involved in the process and less steps have to be taken, and a process that could have been weeks to get access gets potentially reduced to seconds.

[3] https://app.diagrams.net/

6. Validation

The validation consists of two parts. Firstly, we validate using ChatGPT [31], since it should be able to understand the data and compare the scenario with the designed RDF description. This is done by checking if the semantics of the RDF description fit the requirements of the scenario. Then, a survey is conducted with stakeholders involved in a real FDT project. They validated the scenarios and also the semantics of the RDF descriptions.

ChatGPT

ChatGPT, based on OpenAI's GPT-4 architecture, is an advanced language model utilising deep learning to generate human-like text. It excels in text generation, question answering, language translation, and summarization by processing a vast array of topics from its extensive training data. It is not particularly designed to process linked data, but does a sufficient job at this.

The validation by ChatGPT is performed as follows: first, the ODRL offer and request are inserted and ChatGPT is prompted to interpret the offer and request and explain what they mean. Then ChatGPT is asked if an agreement should be made based on the offer and request, after which the agreement is entered and ChatGPT is asked if the agreement is corresponding to the offer and request. Lastly, a description of the scenario is inserted into ChatGPT, and ChatGPT is asked if the before entered RDF matches the scenario description. The full conversations with ChatGPT for each of the scenarios can be found in Appendix B. One thing to note is that the copied RDFs do not fully appear in the chatbox, however this appears to be a visual glitch, since ChatGPT does explain every element of the RDF in the conversation.

In the end, the RDF descriptions of all scenarios of Chapter 4 were validated by ChatGPT, since the explanations of ChatGPT matched the intentions of the designs. The conclusions were especially interesting to read since it captures the essence. Furthermore, the prompts have been entered into multiple different chats, to test if the results would be similar. Even though some of the wording might differ, the general message would be similar. Sometimes ChatGPT made some errors where it said that something was wrong like the policyUsage part from the scenario of Section 4.2. This is an error of ChatGPT, since it tries to replace it with a term that does not exist. These errors were investigated and eventually ignored.

Survey

We also performed a survey to validate the scenarios of Chapter 4, as well as the semantics of the RDFs from the agreements. The survey has been created using Google Forms[4]. The survey has been sent to experts on the FDT that are involved in a real life project involving the FDT conducted by the University of Twente discussed in Section 4.8. The survey can be found using the following URL: https://forms.gle/1riRUGsAGqsmGPea8.

First of all, the respondents must agree to the term of consent. In the term of consent it clearly states the description of the study, what participation in the study involves and how data is collected and handled. Contact details are also provided in case there are questions

[4] https://www.google.nl/intl/nl/forms/about/

or concerns regarding the survey. The respondents have to agree to the conditions that they have read and understand all of this information, they voluntarily agree to participate in the online study, they are 18 years of age or older and that they understand that they can withdraw from participation at all times without a penalty.

After the term of consent has been accepted by the respondents, they are asked if they are able to interpret RDFs. A respondent must be able to do this, since otherwise they are not qualified to validate the RDFs of the agreements. The respondents can simply answer the question with ‘yes’ or ‘no’. If they answer ‘yes’ the survey continues, however, if they answer ‘no’ the survey ends.

Then the respondents are asked if they have ever had to deal with data access conditions. They can answer either ‘yes’ or ‘no’. In case the respondents answer ‘yes’ they move on to the rest of the survey. However, if the respondents answer ‘no’ the survey ends for them, since such a respondent is not qualified to validate the scenarios.

Afterwards, the respondents are asked their role when dealing with data access conditions. They can answer either a data owner, a data access requester, both of the latter options or they can write an answer that is not in the list. Ideally, as many data owners as data access requesters answer the survey to get balanced results. Furthermore, the respondents are asked if they were involved in the project from the University of Twente that involves the FDT, described in the scenario of Section 4.8. Since validation is necessary for this scenario to check whether the essence and the problems of the project have been described properly in the scenario. Then the respondents have to rate their familiarity with linked data or RDF to check their experience with these technologies. The respondents can select an answer from ‘not at all familiar’, ‘slightly familiar’, ‘moderately familiar’, ‘very familiar’ and ‘extremely familiar’. If the respondents answer higher on this scale the results will be more reliable. The same applies to the questions after that, where the respondents rate their familiarity with the FDT/PHT, they can select from the same options of answers. Furthermore, the respondents are asked about their familiarity with ODRL, where the same principle applies as the previous two questions.

The survey asks for validation on the scenarios of Sections 4.7, 4.8 and 4.9, since these scenarios are the most complex and have the most complex RDFs that represent a real life use case. It does so by asking the respondents if they agree that the scenario is realistic and represents a potential real-life situation. They can then answer if they strongly agree, agree, neither agree nor disagree, disagree or strongly disagree with this statement. Furthermore, the respondents are asked if they have ever experienced a situation similar to the scenario, where they can answer ‘yes’ or ‘no’. Then the RDF belonging to the scenario is presented that shows the agreement. Only the agreement is shown, since this is the part that needs validation and includes the most important elements from the access request and access conditions. The respondents are asked if the RDF represents an agreement based on the scenario and are provided with a description of the semantics of the RDF as well as the RDF itself. The description is provided to make it easier to understand the RDF. The respondents can answer if they strongly agree, agree, neither agree nor disagree, disagree or strongly disagree. The respondents are asked additional questions specifically on the scenario of Section 4.8, since it is based on a real project. They are asked if the scenario encapsulates the problem of conflict of interest between the farmers and researchers, and if the duties and constraints from the agreement are useful to deal with this problem.

Scenarios of Sections 4.1 up until 4.6 are not included in the survey, since most of the elements included in the RDFs of these scenarios are included in the scenarios that are validated. Furthermore, Section 4.10 is not validated since it is not really a scenario and it would make the survey too complex and take up too much time from the respondents.

Survey results

The respondent's answers of the survey can be found in Appendix C. In the end, there was only one respondent who answered the survey even though many more were approached. Therefore the survey results are not representative, but the survey questions are useful for future validation. The answers to the first questions determine that the respondent is indeed an expert in the field and qualified to validate the scenarios and RDF descriptions.

Regarding the realism of the scenarios, the respondent has indicated that they are neutral about the scenario from Section 4.7 and agree that the scenarios from Sections 4.8 and 4.9 are realistic. Furthermore, the respondent has indicated that they have experienced similar scenarios in real life before, further indicating that the scenarios have realistic aspects. The respondent has answered that they agree that the RDF descriptions correctly represent the scenarios. Lastly, the respondent has indicated that they are neutral about the constraints and duties being sufficient to reduce data sharing reluctancy from the farmers.

Because of the mostly positive feedback from the respondent no reconsiderations have to be made on the scenarios and the corresponding RDF descriptions. However, in the future, another method of qualitative research, e.g. interviews, could be performed to indicate why the respondent did not fully agree on certain parts.

7. Conclusion

In conclusion, this thesis has explored the potential of automating machine-readable data access agreements using ODRL within the framework of a FDT. The research addressed the critical need for efficient and secure data sharing in an era of rapidly increasing data generation and increasing regulatory requirements. By developing a standardised methodology for creating and validating ODRL-based agreements, the study provides a robust solution to facilitate automated data access while ensuring compliance with data ownership and privacy regulations.

The main research question that this thesis was aimed to answer was: How can ODRL be used to support data access authorization in federated analysis platforms such as the FDT?

The results from Chapter 4 show that there is a common method to create ODRL agreements based on the ODRL offer and ODRL request. The request should match the permissions and prohibitions of the offer for an agreement to be made possible. Then these matching permissions and prohibitions are all included in the ODRL agreement between the data owner and researcher. The permissions, prohibitions and target asset can always be deduced from the offer, these form the basis for the agreement.

For this process to be automated, the algorithm described in Chapter 5 should be implemented. The algorithm describes what steps should be taken before an agreement can be reached between a researcher and data owner. Combining the knowledge of Chapters 4 and 5, the main research question can be answered. This results in an algorithm that runs on the data owners side and compares the access request with access conditions. If there is a match, the ODRL agreement should be created by using the ODRL offer with its target asset, permissions and prohibitions as a basis and adding the dcterms:references to refer to the request. Then the assignee can be added and an agreement is made and access can be granted to the researcher. To further speed up the process, a standardised machine-readable access request is designed that should comply with most access conditions. An example of a standardised access request is described in Section 4.10.

In summary, this thesis contributes to the advancement of automated data access management, offering a viable approach to streamline data sharing processes in compliance with FAIR principles and regulatory standards. The proposed solutions pave the way for more efficient and secure data access, fostering innovation and collaboration in various fields reliant on data-driven insights.

7.1 Discussion

This thesis focuses on data access automation, however it begs the question: is automation necessarily always a good improvement? Up to an extent, automation will be a necessity in the near future, due to the increasing amount of data and is therefore inevitable. However, in regard to very sensitive data, we think that some form of human monitoring and control will have to be involved. This way, misuse of the automation can be prevented and a higher level of security can be reached for data. However, we think that the number of cases where human involvement is necessary should be minimised to increase efficiency.

There could even be a way where both automation and human involvement can be combined. Instead of automatically denying access to a researcher when conditions do not match, a human could check the differences between the conditions and request manually. The differences should then be pointed out by the algorithm. This would already improve efficiency, since a human only has to check the differences and not the matching parts, instead of comparing the entire request with the conditions. In this case, whenever an agreement is made, the duties and rules of the request should be used as a basis for the agreement instead of the usuals rules and duties from the offer.

Furthermore, before the process can be fully automated, there should be a method to check whether a party has breached the agreement. This could be performed by a third-party application. Once the agreement has been breached, a penalty should be applied or legal action can be taken. Agreements must be enforced for data owners, otherwise the automation technology will likely not be applied since their data might be used in a different way than intended.

Lastly, for full automation to be applicable, there has to be some sort of third party tool that validates all of the information provided. For example, if a company only wants their data to be used for academic research, there has to be a form of validation in place to check whether the access request comes from an actual affiliate from a university, instead of somebody pretending to be an academic researcher. This tool should also be able to check if prerequisite conditions are fulfilled before access should be granted to data and should ideally be based on the FAIR principles. Before tools to improve security and authenticity of requests are available, we think it is too early to implement full automation of data access. However, eventually full automation will be a necessity, due to the large amount of data being generated.

7.2 Limitations

The results of this research suffer from several limitations. Firstly, since data access automation is quite a novel field, the research conducted on this topic is limited. Most of the research already conducted is similar to the literature discussed in Section 2.2, however, this is quite limited. Furthermore, since ODRL is currently not widely applied, it does not have a proper guide or tutorial of how to apply it. Learning ODRL took valuable time leaving us less time to develop the algorithm, which fell outside of the scope of the thesis. Moreover, the validation of the research is limited, since the number of respondents is too small. Not many researchers are experts in this domain and therefore the target group is still small. However, we approached more researchers. But they did not respond and cannot be forced to. Lastly, since a survey was used for validation, the respondent did not have a way of providing detailed feedback. They indicate that they did not fully agree with some statements but were unable to provide a reason as to why.

7.3 Future work

This thesis contributes to the field of data access automation by providing a basis for the creation of ODRL agreements based on an access request and an access policy. However, there is still much work to be done in the future. The algorithm described in Chapter 5 should be fully implemented. Ideally, in the future, some third party tools should be able to connect with this algorithm as well, to provide authentication, for example. Furthermore, once the algorithm is developed, it should be tested in an FDT by adding the access request into the metadata of the train. The data stations should then also have the ODRL offers added as metadata. These tests should be performed in different scenarios using different kinds of data, like medical data, financial data, etc., with different kinds of data access conditions in place. Additionally, this thesis mostly considers researchers launching an FDT for research purposes, but additional research could be conducted to find out whether there are applications for automating such a process in businesses. Moreover, additional validation will be needed once the automated matching algorithm has been implemented. Lastly, further validation of the current scenarios should be performed to validate the scenarios in realistic situations. This includes an interview that could be conducted with the target group of the survey, since a survey does not give them the option to give detailed answers.

References

[1] Guth, S. (2003). Rights expression languages. In Lecture notes in computer science (pp. 101–112). https://doi.org/10.1007/10941270_8

[2] Mourtzis, D. (2021). Towards the 5th Industrial Revolution: A literature review and a framework for process optimization based on big data analytics and semantics. Journal of Machine Engineering. https://doi.org/10.36897/jme/141834

[3] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., Da Silva Santos, L. O. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T. W., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific Data, 3(1). https://doi.org/10.1038/sdata.2016.18

[4] General Data Protection Regulation (GDPR) – Official Legal text. (2022, 27 September). General Data Protection Regulation (GDPR). https://gdpr-info.eu/

[5] Da Silva Santos, L. O. B., Pires, L. F., Martinez, V. G., Moreira, J., & Guizzardi, R. S. S. (2022). Personal health train architecture with dynamic cloud staging. SN Computer Science, 4(1). https://doi.org/10.1007/s42979-022-01422-4

[6] GÜLBEY, B. G. (2023). Semantic model of the Computing Capacity matching within the FAIR Data Train. TScIT 39. https://essay.utwente.nl/95815/1/G%C3%BClbey_BA_EEMCS.pdf

[7] ODRL Information Model 2.2. (2018, 15 February). https://www.w3.org/TR/odrl-model/

[8] Cheah, P. Y., & Piasecki, J. (2020). Data access committees. BMC Medical Ethics, 21(1). https://doi.org/10.1186/s12910-020-0453-z

[9] Informed Consent and ethics committees. Research Data Management. https://www.ru.nl/rdm/collecting-data/informed-consent-ethics-committees/#hea6e9410-b04e-a843-6113-d808a16cf128

[10] Jamkhedkar, P. A., & Heileman, G. L. (2009). Chapter 1. In Rights Expression Languages. https://ece-research.unm.edu/informatics/publications/Rights%20Expression%20Lanagues.pdf

[11] Coyle, K. (2004, February). Rights Expression Languages. https://www.loc.gov/standards/relreport.pdf

[12] Homepage - Creative Commons. (2023, 16 November). Creative Commons. https://creativecommons.org/

[13] Draft rights Declaration schema is ready for review: Metadata Encoding and Transmission Standard (METS) OfficialWeb site.https://www.loc.gov/standards/mets/news080503.html

[14] Standards – MPEG. https://www.mpeg.org/standards/MPEG-21/

[15] EXtensible Access Control Markup Language (XACML) version 3.0. https://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html

[16] Ramli, C. D. P. K., Nielson, H. R., & Nielson, F. (2014). The logic of XACML. Science of Computer Programming, 83, 80–105. https://doi.org/10.1016/j.scico.2013.05.003

[17] Pellegrini, T., Havur, G., Steyskal, S., Panasiuk, O., Fensel, A., Fachhochschule St. Pölten, Department Medienwirtschaft, Mireles, V., Thurner, T., Vienna University of Economics and Business, Department of Informations Systems & Operations, Polleres, A., Kirrane, S., STI Innsbruck, University of Innsbruck, & The Semantic Web Company. (2018). DALICC: a LICENSE MANAGEMENT FRAMEWORK FOR DIGITAL ASSETS [Journal-article]. DALICC: A LICENSE MANAGEMENT FRAMEWORK FOR DIGITAL ASSETS, 1. https://penni.wu.ac.at/papers/IRIS%202019%20DALICC%20A%20License%20Management%20Framework%20for%20Digital%20Assets.pdf

[18] DALICC – Data Licenses Clearance Center – license management made easy. https://www.dalicc.net/

[19] Lawson, J., Cabili, M. N., Kerry, G., Boughtwood, T., Thorogood, A., Alper, P., Bowers, S. R., Boyles, R. R., Brookes, A. J., Brush, M., Burdett, T., Clissold, H., Donnelly, S., Dyke, S. O., Freeberg, M. A., Haendel, M. A., Hata, C., Holub, P., Jeanson, F., . . . Courtot, M. (2021). The Data Use Ontology to streamline responsible access to human biomedical datasets. Cell Genomics, 1(2), 100028. https://doi.org/10.1016/j.xgen.2021.100028

[20] Data Use Ontology (DUO). https://www.ga4gh.org/product/data-use-ontology-duo/

[21] Wieringa, R. J. (2014). Design Science Methodology for Information Systems and Software Engineering. In Springer eBooks. https://doi.org/10.1007/978-3-662-43839-8

[22] ODRL Vocabulary & Expression 2.2. (2018, 15 February). https://www.w3.org/TR/odrl-vocab/

[23] Beveridge, A., Hansen, J. B., Val, J., Gehrmann, L., Farmer, R., Khutan, S., Robertson, T., & Llamas, M. Validata: RDF Validator. https://www.w3.org/2015/03/ShExValidata/

[24] isSemantic.net. isSemantic.net - structured data tool. https://issemantic.net/rdf-visualizer

[25] ISO - ISO 3166 — Country Codes. ISO. https://www.iso.org/iso-3166-country-codes.html

[26] Data Privacy Vocabulary (DPV). (2024, 1 January). https://w3c.github.io/dpv/dpv/

[27] Shapes Constraint Language (SHACL). (2017, 20 July). https://www.w3.org/TR/shacl/

[28] DCMI Metadata Terms.
https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

[29] Del Carmen Sanchez Gonzalez, M., Kamerling, P., Iermito, M., Casati, S., Riaz, U., Veal, C. D., Maini, M., Jeanson, F., Benhamed, O. M., Van Enckevort, E., Landi, A., Mimouni, Y., Cornec, C. L., Coviello, D. A., Franchin, T., Fusco, F., García, J. A. R., Van Der Zanden, L. F. M., Bernier, A., . . . Brookes, A. J. (2024). Common conditions of use elements. Atomic concepts for consistent and effective information governance. Scientific Data, 11(1). https://doi.org/10.1038/s41597-024-03279-z

[30] Wilkinsonlab. GitHub - wilkinsonlab/odrl-translator-demo. GitHub.
https://github.com/wilkinsonlab/odrl-translator-demo/tree/main

[31] OpenAI. (2023). GPT-4 [Large language model]. OpenAI. https://www.openai.com

[32] Rachel L Richesson, Prakash Nadkarni, Data standards for clinical research data collection forms: current status and challenges, Journal of the American Medical Informatics Association, Volume 18, Issue 3, May 2011, Pages 341–346,
https://doi.org/10.1136/amiajnl-2011-000107

Appendix

A RDFs of ODRL Agreements based on scenarios

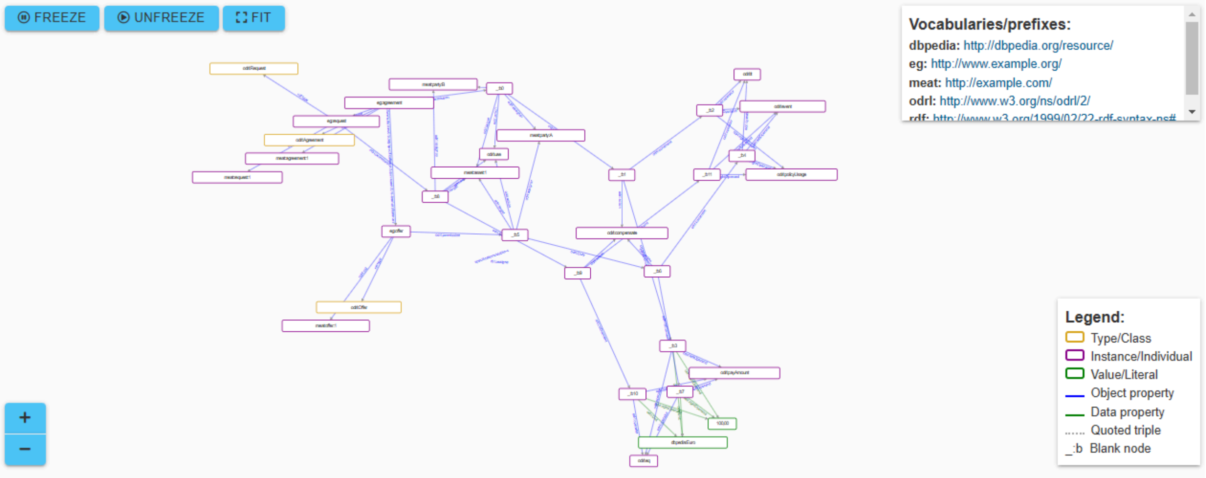
A.1 Compensation involved

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .

ex:offer a odrl:Offer ;
        odrl:uid <http://example.com/offer:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:action odrl:use ;
                odrl:Duty [
                        odrl:action odrl:compensate ;
                                odrl:refinement [
                                odrl:leftOperand odrl:payAmount ;
                                odrl:operator odrl:eq ;
                                odrl:rightOperand "100,00"^^xsd:decimal ;
                                odrl:Unit "http://dbpedia.org/resource/Euro" ;
                        ] ;
                        odrl:constraint [
                                odrl:leftOperand odrl:event ;
                                odrl:operator odrl:lt ;
                                odrl:rightOperand odrl:policyUsage ;
                        ] ;
                ] ;
        ] .

ex:request a odrl:Request ;
        odrl:uid <http://example.com/request:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:use ;
                odrl:Duty [
                        odrl:action odrl:compensate ;
                                odrl:refinement [
                                odrl:leftOperand odrl:payAmount ;
                                odrl:operator odrl:eq ;
                odrl:rightOperand "100,00"^^xsd:decimal ;
                odrl:Unit "http://dbpedia.org/resource/Euro" ;
            ] ;
            odrl:constraint [
                odrl:leftOperand odrl:event ;
                odrl:operator odrl:lt ;
                odrl:rightOperand odrl:policyUsage ;
            ] ;
        ] ;
    ] .

ex:agreement a odrl:Agreement ;
    dcterms:references ex:offer, ex:request ;
    odrl:uid <http://example.com/agreement:1> ;
    odrl:permission [
        odrl:target <http://example.com/asset:1> ;
        odrl:assigner <http://example.com/party:A> ;
        odrl:assignee <http://example.com/party:B> ;
        odrl:action odrl:use ;
        odrl:Duty [
            odrl:action odrl:compensate ;
                odrl:refinement [
                odrl:leftOperand odrl:payAmount ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "100,00"^^xsd:decimal ;
                odrl:Unit "http://dbpedia.org/resource/Euro" ;
            ] ;
            odrl:constraint [
                odrl:leftOperand odrl:event ;
                odrl:operator odrl:lt ;
                odrl:rightOperand odrl:policyUsage ;
            ] ;
        ] ;
    ] .
```



Vocabularies/prefixes:
dbpedia: http://dbpedia.org/resource/
eg: http://www.example.org/
meat: http://example.com/
odrl: http://www.w3.org/ns/odrl/2/
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#

Figure A.1: The triples of RDF A1.2.

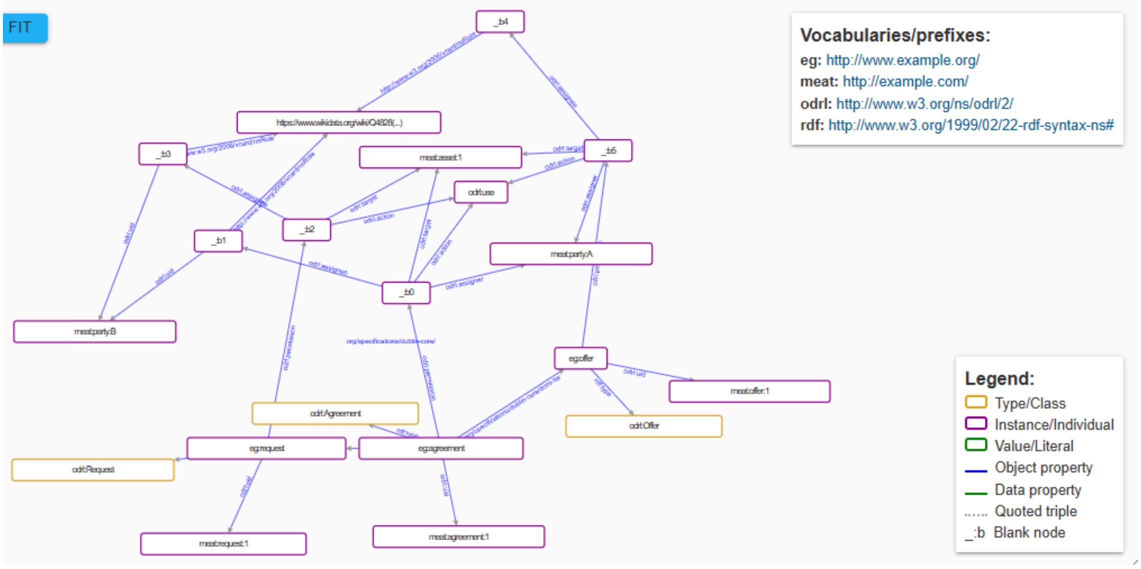
A.2 Roles involved.

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .

ex:offer a odrl:Offer ;
        odrl:uid <http://example.com/offer:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:assignee [
                        vcard:Role <https://www.wikidata.org/wiki/Q48282> ;
                ] ;
                odrl:action odrl:use ;
        ] .

ex:request a odrl:Request ;
        odrl:uid <http://example.com/request:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assignee [
                        odrl:uid <http://example.com/party:B> ;
                        vcard:Role <https://www.wikidata.org/wiki/Q48282> ;
                ] ;
                odrl:action odrl:use ;
        ] .

ex:agreement a odrl:Agreement ;
        dcterms:references ex:offer, ex:request ;
        odrl:uid <http://example.com/agreement:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:assignee [
                        odrl:uid <http://example.com/party:B> ;
                        vcard:Role <https://www.wikidata.org/wiki/Q48282> ;
                ] ;
                odrl:action odrl:use ;
        ] .
```



Vocabularies/prefixes:
eg: http://www.example.org/
meat: http://example.com/
odrl: http://www.w3.org/ns/odrl/2/
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#

Figure A.2: The triples of RDF A1.3.

A.3 The use of a specific action and prohibition

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .

ex:offer a odrl:Offer ;
        odrl:uid <http://example.com/offer:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:action odrl:reproduce ;
        ] ;
        odrl:prohibit [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:action odrl:modify ;
        ] .

ex:request a odrl:Request ;
        odrl:uid <http://example.com/request:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:reproduce ;
        ] ;
        odrl:prohibit [
                odrl:target <http://example.com/asset:1> ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:modify ;
        ] .

ex:agreement a odrl:Agreement ;
        dcterms:references ex:offer, ex:request ;
        odrl:uid <http://example.com/agreement:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:reproduce ;
        ] ;
        odrl:prohibit [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:modify ;
        ] .
```

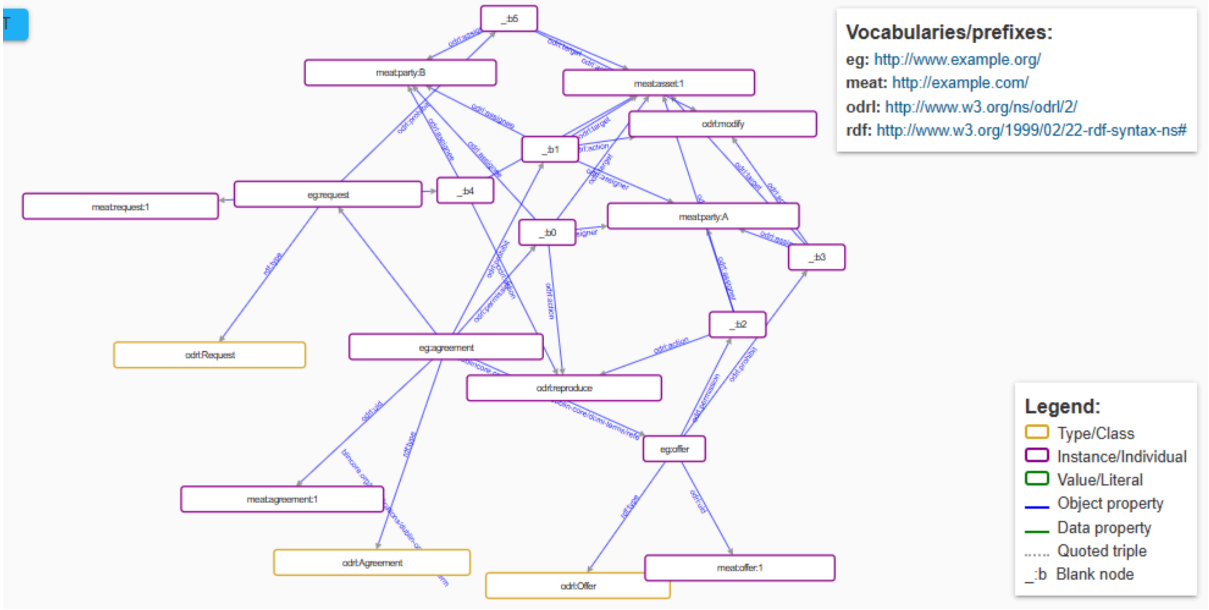


Figure A.3: The triples of RDF A1.4.

A.4 Basic constraints

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
```

```
@prefix vcard: <http://www.w3.org/2006/vcard/ns> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .

ex:offer a odrl:Offer ;
    odrl:uid <http://example.com/offer:1> ;
    odrl:permission [
        odrl:target <http://example.com/asset:1> ;
        odrl:assigner <http://example.com/party:A> ;
        odrl:action odrl:use ;
        odrl:Duty [
            odrl:constraint [
                odrl:leftOperand odrl:spatial ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "NLD" ;
            ] ,
            [
                odrl:leftOperand odrl:dateTime ;
                odrl:operator odrl:lt ;
                odrl:rightOperand "2026-01-01"^^xsd:date ;
            ] ;
        ] ;
    ] .

ex:request a odrl:Request ;
    odrl:uid <http://example.com/request:1> ;
    odrl:permission [
        odrl:target <http://example.com/asset:1> ;
        odrl:assignee <http://example.com/party:B> ;
        odrl:action odrl:use ;
        odrl:Duty [
            odrl:constraint [
                odrl:leftOperand odrl:spatial ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "NLD" ;
            ] ,
            [
                odrl:leftOperand odrl:dateTime ;
                odrl:operator odrl:lt ;
                odrl:rightOperand "2026-01-01"^^xsd:date ;
            ] ;
        ] ;
    ] .


ex:agreement a odrl:Agreement ;
    dcterms:references ex:offer, ex:request ;
    odrl:uid <http://example.com/agreement:1> ;
    odrl:permission [
            odrl:target <http://example.com/asset:1> ;
            odrl:assigner <http://example.com/party:A> ;
            odrl:assignee <http://example.com/party:B> ;
            odrl:action odrl:use ;
            odrl:Duty [
                        odrl:constraint [
                                    odrl:leftOperand odrl:spatial ;
                                    odrl:operator odrl:eq ;
                                    odrl:rightOperand "NLD" ;
                        ] ,
                        [
                                    odrl:leftOperand odrl:dateTime ;
                                    odrl:operator odrl:lt ;
                                    odrl:rightOperand "2026-01-01"^^xsd:date ;
                        ] ;
            ] ;
] .
```

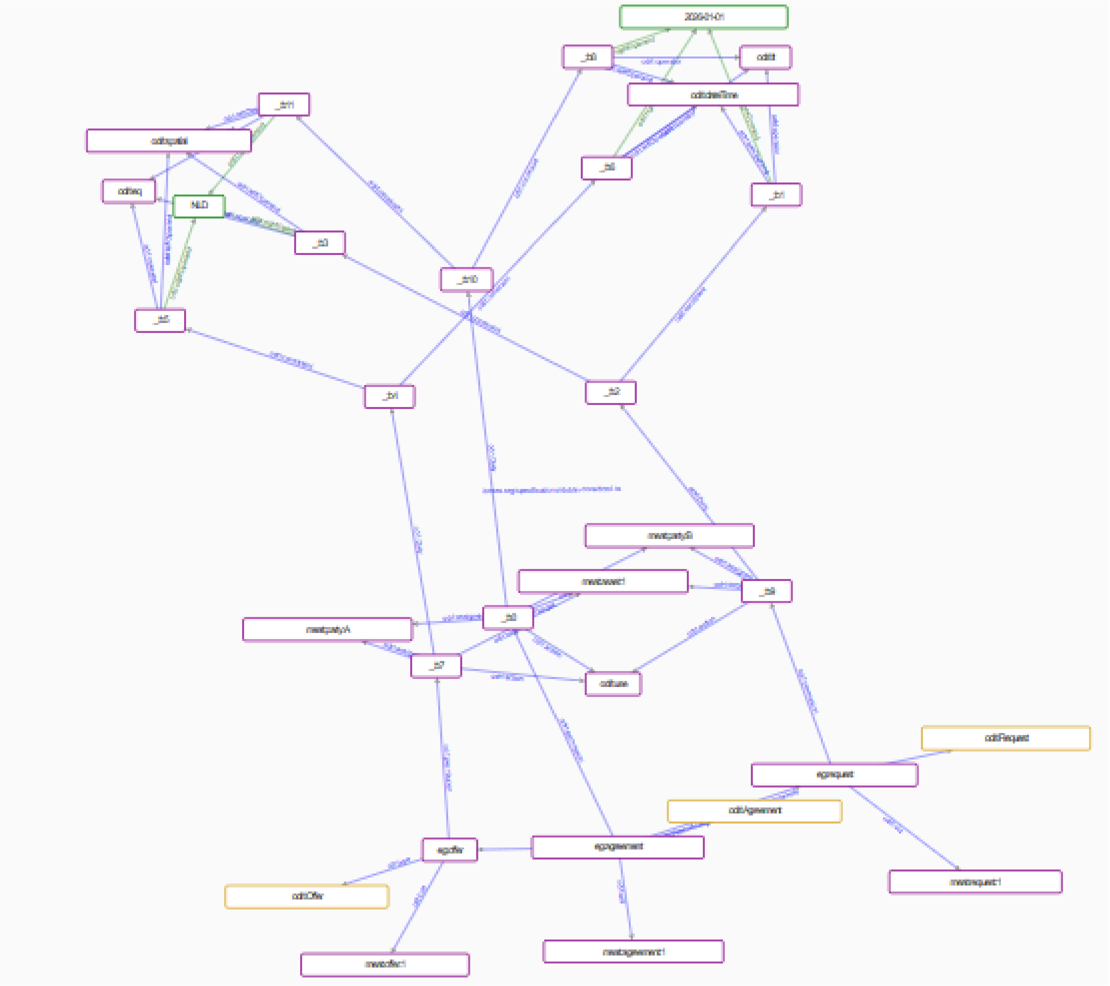


Figure A.4: The triples of RDF A1.5.

A.5 Mismatch example

@prefix odrl: <http://www.w3.org/ns/odrl/2/> .

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .

ex:offer a odrl:Offer ;
        odrl:uid <http://example.com/offer:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:action odrl:use ;
                odrl:Duty [
                        odrl:constraint [
                                odrl:leftOperand odrl:spatial ;
                                odrl:operator odrl:eq ;
                                odrl:rightOperand "NLD" ;
                        ] ,
                        [
                                odrl:leftOperand odrl:dateTime ;
                                odrl:operator odrl:lt ;
                                odrl:rightOperand "2026-01-01"^^xsd:date ;
                        ] ;
                ] ;
        ] .

ex:request a odrl:Request ;
        odrl:uid <http://example.com/request:1> ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:use ;
                odrl:Duty [
                        odrl:constraint [
                                odrl:leftOperand odrl:spatial ;
                                odrl:operator odrl:eq ;
                                odrl:rightOperand "BRA" ;
                        ] ,
                        [
                                odrl:leftOperand odrl:dateTime ;
                                odrl:operator odrl:lt ;
                                odrl:rightOperand "2027-01-01"^^xsd:date ;
                        ] ;
                ] ;
        ] .
```

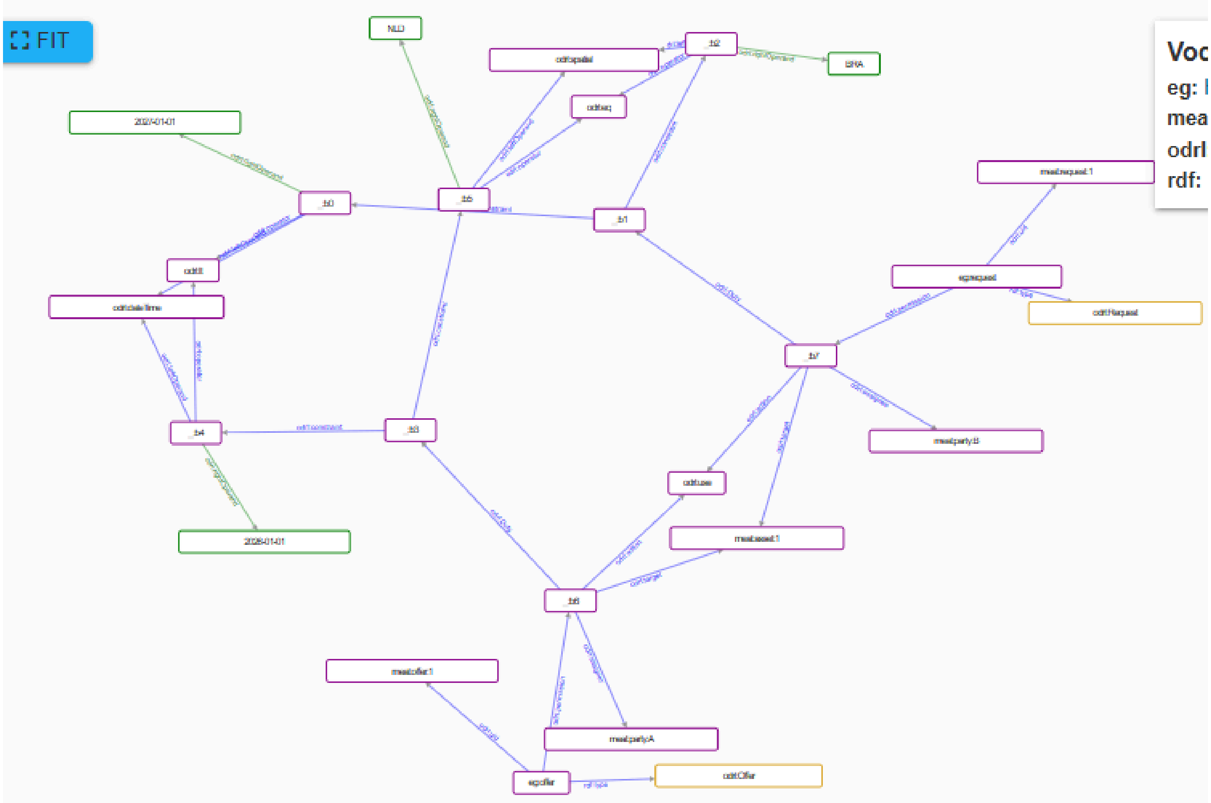


Figure A.5: The triples of RDF A1.6.

A.6 Combination of former scenarios

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .
@prefix vcard: <https://www.w3.org/TR/vcard-rdf/> .

        ex:offer a odrl:Offer ;
          odrl:uid <http://example.com/offer:1> ;
          odrl:permission [
            odrl:target <http://example.com/asset:1> ;
            odrl:assigner <http://example.com/party:A> ;
            odrl:assignee [
              vcard:Role <https://www.wikidata.org/wiki/Q1650915> ;
            ] ;
            odrl:action odrl:use ;
            odrl:Duty [
              odrl:action odrl:compensate ;
              odrl:refinement [
                odrl:leftOperand odrl:payAmount ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "300.00"^^xsd:decimal ;
                odrl:unit <http://dbpedia.org/resource/United_States_dollar> ;
              ] ;
              odrl:constraint [
                odrl:leftOperand odrl:event ;
                odrl:operator odrl:lt ;
                odrl:rightOperand odrl:policyUsage ;
              ] ;
            ] ;
            odrl:Duty [
              odrl:action odrl:anonymize ;
            ] ;
          ] .


        ex:request a odrl:Request ;
              odrl:uid <http://example.com/request:1> ;
              odrl:permission [
                    odrl:target <http://example.com/asset:1> ;
                    odrl:assignee [
                          odrl:uid <http://example.com/party:B> ;
                          vcard:Role <https://www.wikidata.org/wiki/Q1650915> ;
                    ] ;
                    odrl:action odrl:use ;
                    odrl:Duty [
                    odrl:action odrl:compensate ;
                    odrl:refinement [
                          odrl:leftOperand odrl:payAmount ;
                          odrl:operator odrl:eq ;
                          odrl:rightOperand "300,00"^^xsd:decimal ;
                          odrl:Unit "http://dbpedia.org/resource/United_States_dollar" ;
                          ] ;
                    odrl:constraint [
                          odrl:leftOperand odrl:event ;
                          odrl:operator odrl:lt ;
                          odrl:rightOperand odrl:policyUsage ;
                          ] ;
                    ] ;
                    odrl:Duty [
                          odrl:action odrl:anonymize ;
                          ] ;
                    ] .

        ex:agreement a odrl:Agreement ;
              dcterms:references ex:offer, ex:request ;
              odrl:uid <http://example.com/agreement:1> ;
              odrl:permission [
                    odrl:target <http://example.com/asset:1> ;
                    odrl:assigner <http://example.com/party:A> ;
            odrl:assignee [
                odrl:uid <http://example.com/party:B> ;
                vcard:Role <https://www.wikidata.org/wiki/Q1650915> ;
            ] ;
            odrl:action odrl:use ;
            odrl:Duty [
            odrl:action odrl:compensate ;
            odrl:refinement [
                odrl:leftOperand odrl:payAmount ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "300,00"^^xsd:decimal ;
                odrl:Unit "http://dbpedia.org/resource/United_States_dollar" ;
                ] ;
            odrl:constraint [
                odrl:leftOperand odrl:event ;
                odrl:operator odrl:lt ;
                odrl:rightOperand odrl:policyUsage ;
                ] ;
            ] ;
            odrl:Duty [
                odrl:action odrl:anonymize ;
            ] ;
        ] .
```

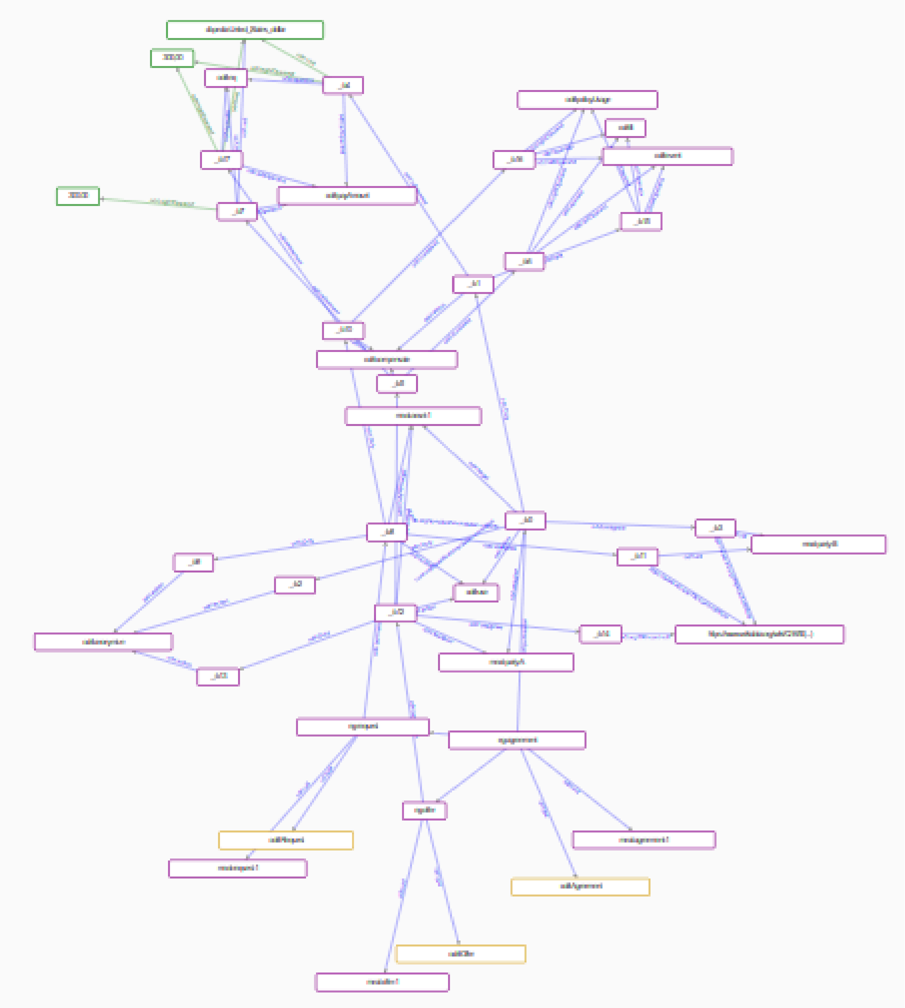


Figure A.6: The triples of RDF A1.7.

A.7 A real life FDT use case

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .
@prefix vcard: <https://www.w3.org/TR/vcard-rdf/> .

ex:offer a odrl:Offer ;
        odrl:uid <http://example.com/offer:1> ;
        odrl:permission [
                odrl:action odrl:use ;
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <https://www.wikidata.org/wiki/Q131512>;
                odrl:Duty [
                        odrl:constraint [
                        odrl:leftOperand odrl:purpose ;
                odrl:operator odrl:eq ;
                odrl:rightOperand dpv:CommercialPurpose ;
            ] ;
            odrl:action dpv:SellProducts ;
            ] ;
        odrl:Duty [
            odrl:action dpv:Marketing ;
            ];
        odrl:Duty [
            odrl:action odrl:inform ;
            ] ;
        odrl:Duty [
            odrl:action <http://creativecommons.org/ns#SourceCode> ;
            ] ;
        odrl:Duty [
            odrl:action odrl:Include ;
            odrl:refinement [
                odrl:leftOperand odrl:media;
                odrl:operator odrl:eq;
                odrl:rightOperand dpv:IdentityAuthentication ;
            ] ;
            odrl:constraint [
                odrl:leftOperand odrl:event ;
                odrl:operator odrl:lt ;
                odrl:rightOperand odrl:policyUsage ;
            ] ;

        ] ;
    ].

ex:request a odrl:Request ;
    odrl:uid <http://example.com/request:1> ;
    odrl:permission [
        odrl:action odrl:use ;
        odrl:target <http://example.com/asset:1> ;
        odrl:assignee <http://example.com/party:B>;
        odrl:Duty [
            odrl:constraint [
            odrl:leftOperand odrl:purpose ;
                odrl:operator odrl:eq ;
                odrl:rightOperand dpv:CommercialPurpose ;
            ] ;
            odrl:action dpv:SellProducts ;
            ] ;
        odrl:Duty [
            odrl:action dpv:Marketing ;
        ];
        odrl:Duty [
            odrl:action odrl:inform ;
            ] ;
        odrl:Duty [
            odrl:action <http://creativecommons.org/ns#SourceCode> ;
            ] ;
        odrl:Duty [
            odrl:action odrl:Include ;
            odrl:refinement [
                odrl:leftOperand odrl:media;
                odrl:operator odrl:eq;
                odrl:rightOperand dpv:IdentityAuthentication ;
            ] ;
            odrl:constraint [
                odrl:leftOperand odrl:event ;
                odrl:operator odrl:lt ;
                odrl:rightOperand odrl:policyUsage ;
            ] ;

        ] ;
    ].


ex:agreement a odrl:Agreement ;
dcterms:references ex:offer, ex:request ;
odrl:uid <http://example.com/agreement:1> ;
    odrl:permission [
        odrl:action odrl:use ;
        odrl:target <http://example.com/asset:1> ;
        odrl:assigner <https://www.wikidata.org/wiki/Q131512> ;
        odrl:assignee <http://example.com/party:B>;
        odrl:Duty [
            odrl:constraint [
            odrl:leftOperand odrl:purpose ;
                odrl:operator odrl:eq ;
                odrl:rightOperand dpv:CommercialPurpose ;
            ] ;
            odrl:action dpv:SellProducts ;
            ] ;
        odrl:Duty [
            odrl:action dpv:Marketing ;
            ];
        odrl:Duty [
            odrl:action odrl:inform ;
            ] ;
        odrl:Duty [
            odrl:action <http://creativecommons.org/ns#SourceCode> ;
            ] ;
            odrl:Duty [
                        odrl:action odrl:Include ;
                        odrl:refinement [
                                    odrl:leftOperand odrl:media;
                                    odrl:operator odrl:eq;
                                    odrl:rightOperand dpv:IdentityAuthentication ;
                        ] ;
                        odrl:constraint [
                                    odrl:leftOperand odrl:event ;
                                    odrl:operator odrl:lt ;
                                    odrl:rightOperand odrl:policyUsage ;
                        ] ;

            ] ;
].
```

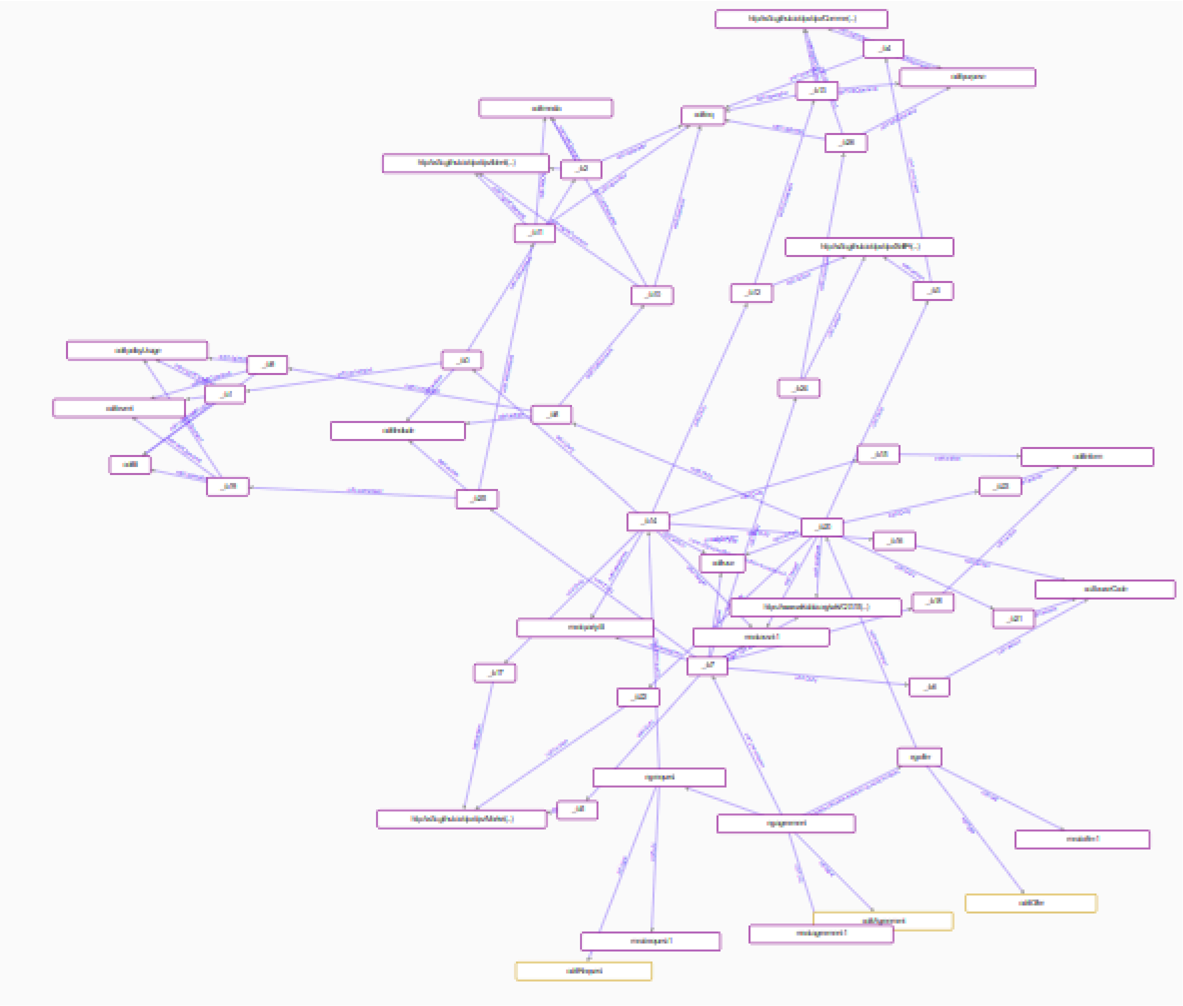


Figure A.7: The triples of RDF A1.8.

A.8 A non-specified asset in the request

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
```

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix wd: <https://www.wikidata.org/wiki/> .
@prefix sh: <http://www.w3.org/ns/shacl#> .

ex:offer a odrl:Offer ;
    odrl:uid <http://example.com/offer:1> ;
    dc:description "This rdf contains personal health data about people and their
    respective blood types. It also has information containing if the person with the blood
    type suffers from any diseases." ;
    dc:issued "2020-01-01T12:00" ;
    odrl:permission [
        odrl:target <http://example.com/asset:1> ;
        odrl:assigner <http://example.com/party:A> ;
        odrl:action odrl:use ;
        odrl:Duty [
            odrl:action odrl:Include ;
            odrl:refinement [
                odrl:leftOperand odrl:media;
                odrl:operator odrl:eq;
                odrl:rightOperand dpv:IdentityAuthentication ;
            ] ;
            odrl:constraint [
            odrl:leftOperand odrl:event ;
            odrl:operator odrl:lt ;
            odrl:rightOperand odrl:policyUsage ;
            ] ;
        ] ;
        odrl:Duty [
            odrl:action odrl:anonymize ;
            ] ;
        ] .

ex:request a odrl:Request ;
    odrl:uid <http://example.com/request:1> ;
    odrl:permission [
        odrl:target [
            a sh:NodeShape ;
            sh:targetClass odrl:Asset ;
            sh:property [
                sh:path dcterms:conformsTo ;
                sh:hasValue wd:Q7873 ;
            ] ;
```

```
                ] ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:use ;
                odrl:Duty [
                        odrl:action odrl:Include ;
                        odrl:refinement [
                                odrl:leftOperand odrl:media;
                                odrl:operator odrl:eq;
                                odrl:rightOperand dpv:IdentityAuthentication ;
                        ] ;
                        odrl:constraint [
                                odrl:leftOperand odrl:event ;
                                odrl:operator odrl:lt ;
                                odrl:rightOperand odrl:policyUsage ;
                        ] ;
                ] ;
                odrl:Duty [
                        odrl:action odrl:anonymize ;
                ] ;
        ] .

ex:agreement a odrl:Agreement ;
        dcterms:references ex:offer, ex:request ;
        odrl:uid <http://example.com/agreement:1> ;
        dc:description "This rdf contains personal health data about people and their
        respective blood types. It also has information containing if the person with the blood
        type suffers from any diseases." ;
        dc:issued "2020-01-01T12:00" ;
        odrl:permission [
                odrl:target <http://example.com/asset:1> ;
                odrl:assigner <http://example.com/party:A> ;
                odrl:assignee <http://example.com/party:B> ;
                odrl:action odrl:use ;
                odrl:Duty [
                        odrl:action odrl:Include ;
                        odrl:refinement [
                                odrl:leftOperand odrl:media;
                                odrl:operator odrl:eq;
                                odrl:rightOperand dpv:IdentityAuthentication ;
                        ] ;
                        odrl:constraint [
                                odrl:leftOperand odrl:event ;
                                odrl:operator odrl:lt ;
                                odrl:rightOperand odrl:policyUsage ;
                        ] ;
                ] ;
                odrl:Duty [
                        odrl:action odrl:anonymize ;
        ] ;
    ] .
```

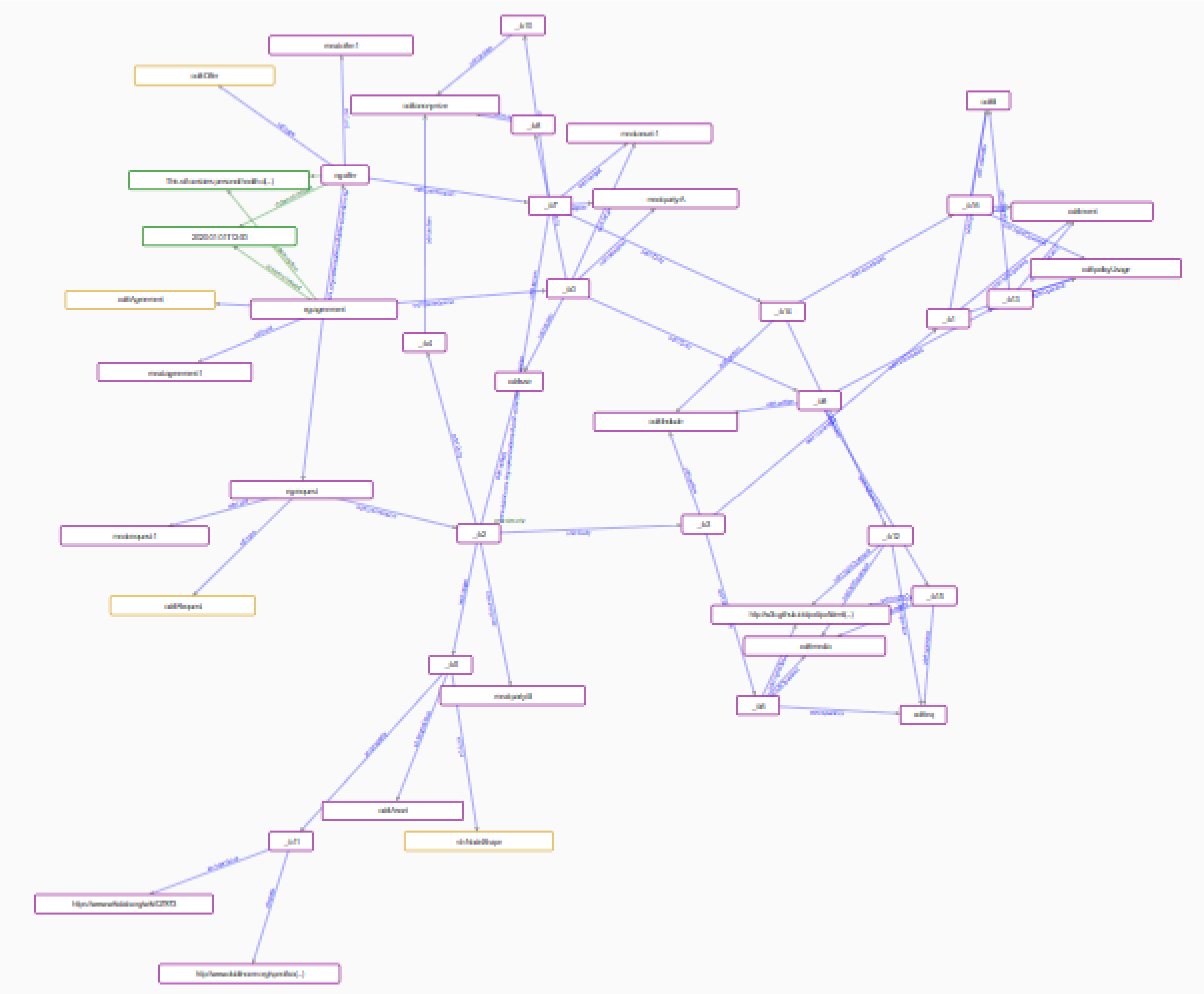


Figure A.8: The triples of RDF A1.9.

A.9 Agreement based on the most common access conditions in health data.

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dpv: <http://w3c.github.io/dpv/dpv/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix ex: <http://www.example.org/> .
@prefix dcterms: <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns> .
@prefix dc: <http://purl.org/dc/terms/> .

ex:offer a odrl:Offer ;
    odrl:uid <http://example.com/offer:1> ;
    odrl:permission [
        odrl:target <http://example.com/asset:1> ;
        odrl:assigner <http://example.com/party:A> ;
```

```
odrl:action odrl:use ;
odrl:constraint [
                odrl:leftOperand odrl:spatial ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "NLD" ;
        ] ,
        [
                odrl:leftOperand odrl:purpose ;
                odrl:operator odrl:eq ;
                odrl:rightOperand dpv:AcademicResearch ;
        ] ,
        [
                odrl:leftOperand odrl:dateTime ;
                odrl:operator odrl:lt ;
                odrl:rightOperand "2124-01-01"^^xsd:date ;
        ] ,
        [
                odrl:leftOperand odrl:event ;
                odrl:operator odrl:lt ;
                odrl:rightOperand odrl:policyUsage ;
        ] ;
odrl:Duty [
        odrl:action odrl:Include ;
        odrl:refinement [
                odrl:leftOperand odrl:media;
                odrl:operator odrl:eq;
                odrl:rightOperand dpv:IdentityAuthentication ;
                ] ;
        ] ;
odrl:Duty [
        odrl:action odrl:anonymize ;
        ] ;
odrl:Duty [
        odrl:action odrl:compensate ;
                odrl:refinement [
                        odrl:leftOperand odrl:payAmount ;
                        odrl:operator odrl:eq ;
                        odrl:rightOperand "0,00"^^xsd:decimal ;
                        odrl:Unit "http://dbpedia.org/resource/Euro" ;
                ] ;
        ];
odrl:Duty [
        odrl:action odrl:inform ;
        ] ;
odrl:Duty [
        odrl:action "http://creativecommons.org/ns#SourceCode" ;
        ] ;
odrl:Duty [
        odrl:action odrl:sharing;
            odrl:refinement [
                odrl:leftOperand odrl:dateTime;
                odrl:operator odrl:gt ;
                odrl:rightOperand "2024-05-01"^^xsd:date ;
            ];

        ] ;
    odrl:Duty [
        odrl:action odrl:obtainConsent;
        ] ;

] .

ex:request a odrl:Request ;
    odrl:uid <http://example.com/request:1> ;
    odrl:permission [
        odrl:target <http://example.com/asset:1> ;
        odrl:assignee <http://example.com/party:B> ;
        odrl:action odrl:use ;
        odrl:constraint [
                odrl:leftOperand odrl:spatial ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "NLD" ;
            ] ,
            [
                odrl:leftOperand odrl:purpose ;
                odrl:operator odrl:eq ;
                odrl:rightOperand dpv:AcademicResearch ;
            ] ,
            [
                odrl:leftOperand odrl:dateTime ;
                odrl:operator odrl:lt ;
                odrl:rightOperand "2124-01-01"^^xsd:date ;
            ] ,
            [
                odrl:leftOperand odrl:event ;
                odrl:operator odrl:lt ;
                odrl:rightOperand odrl:policyUsage ;
            ] ;
        odrl:Duty [
            odrl:action odrl:Include ;
            odrl:refinement [
                odrl:leftOperand odrl:media;
                odrl:operator odrl:eq;
                odrl:rightOperand dpv:IdentityAuthentication ;
                ] ;
            ] ;
        odrl:Duty [
            odrl:action odrl:anonymize ;
            ] ;
        odrl:Duty [
            odrl:action odrl:compensate ;
                odrl:refinement [
                    odrl:leftOperand odrl:payAmount ;
                    odrl:operator odrl:eq ;
                    odrl:rightOperand "0,00"^^xsd:decimal ;
                    odrl:Unit "http://dbpedia.org/resource/Euro" ;
                ] ;
            ];
        odrl:Duty [
            odrl:action odrl:inform ;
            ] ;
        odrl:Duty [
            odrl:action "http://creativecommons.org/ns#SourceCode" ;
            ] ;
        odrl:Duty [
            odrl:action odrl:sharing;
                odrl:refinement [
                    odrl:leftOperand odrl:dateTime;
                    odrl:operator odrl:gt ;
                    odrl:rightOperand "2024-05-01"^^xsd:date ;
                ];

            ] ;
        odrl:Duty [
            odrl:action odrl:obtainConsent;
            ] ;
    ] .

ex:agreement a odrl:Agreement ;
    dcterms:references ex:offer, ex:request ;
    odrl:uid <http://example.com/agreement:1> ;
    odrl:permission [
        odrl:target <http://example.com/asset:1> ;
        odrl:assigner <http://example.com/party:A> ;
        odrl:assignee <http://example.com/party:B>  ;
        odrl:action odrl:use ;
        odrl:constraint [
                odrl:leftOperand odrl:spatial ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "NLD" ;
            ] ,
            [
                odrl:leftOperand odrl:purpose ;
                odrl:operator odrl:eq ;
            odrl:rightOperand dpv:AcademicResearch ;
        ] ,
        [
            odrl:leftOperand odrl:dateTime ;
            odrl:operator odrl:lt ;
            odrl:rightOperand "2124-01-01"^^xsd:date ;
        ] ,
        [
            odrl:leftOperand odrl:event ;
            odrl:operator odrl:lt ;
            odrl:rightOperand odrl:policyUsage ;
        ] ;
    odrl:Duty [
        odrl:action odrl:Include ;
        odrl:refinement [
            odrl:leftOperand odrl:media;
            odrl:operator odrl:eq;
            odrl:rightOperand dpv:IdentityAuthentication ;
            ] ;
        ] ;
    odrl:Duty [
        odrl:action odrl:anonymize ;
        ] ;
    odrl:Duty [
        odrl:action odrl:compensate ;
            odrl:refinement [
                odrl:leftOperand odrl:payAmount ;
                odrl:operator odrl:eq ;
                odrl:rightOperand "0,00"^^xsd:decimal ;
                odrl:Unit "http://dbpedia.org/resource/Euro" ;
            ] ;
        ];
    odrl:Duty [
        odrl:action odrl:inform ;
        ] ;
    odrl:Duty [
        odrl:action "http://creativecommons.org/ns#SourceCode" ;
        ] ;
    odrl:Duty [
        odrl:action odrl:sharing;
            odrl:refinement [
                odrl:leftOperand odrl:dateTime;
                odrl:operator odrl:gt ;
                odrl:rightOperand "2024-05-01"^^xsd:date ;
            ];

        ] ;
    odrl:Duty [
                    odrl:action odrl:obtainConsent;
                    ] ;
          ].
```

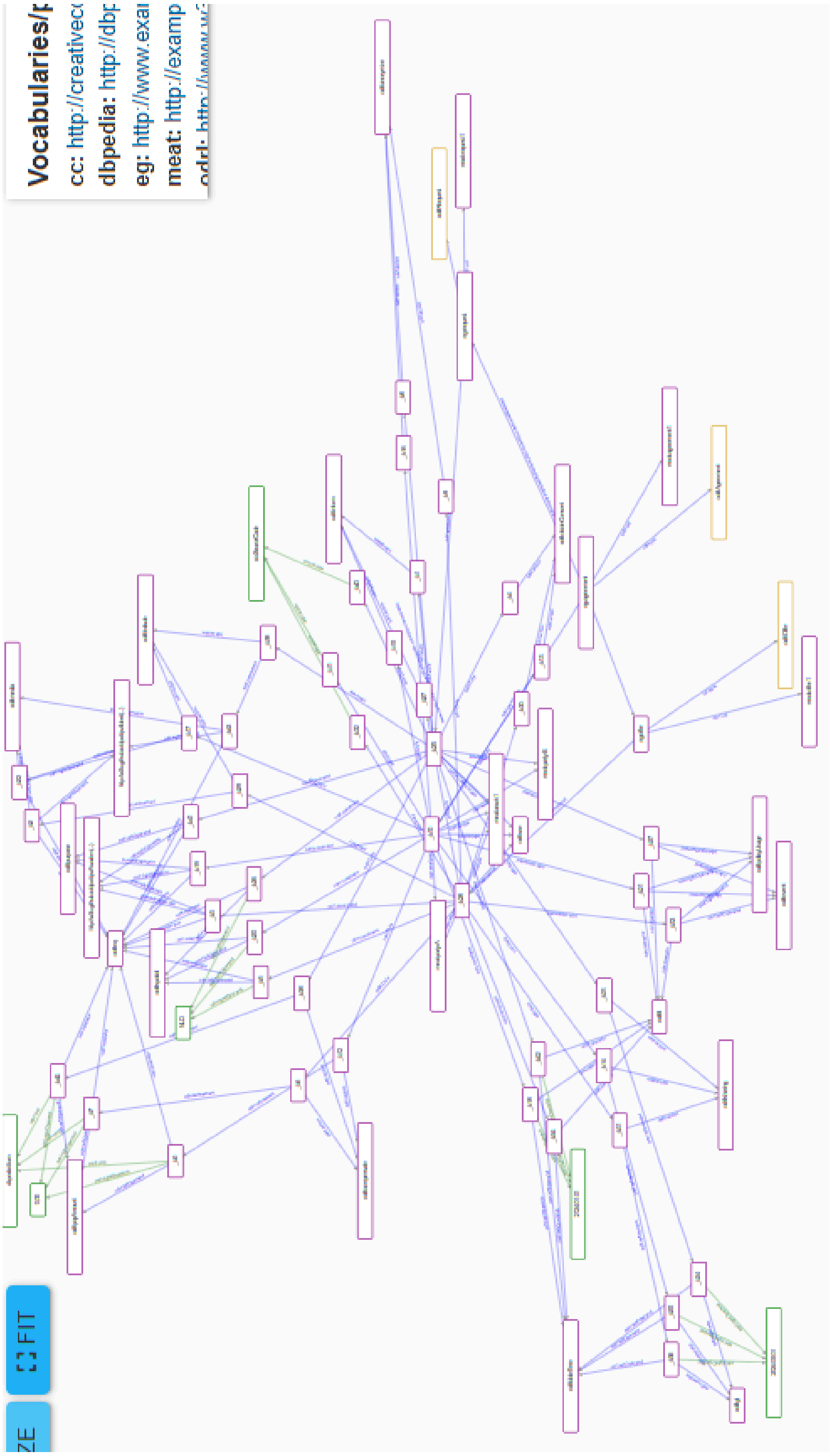


Figure A.9: The triples of RDF A1.10.

B ChatGPT validation responses

A2.1 An open policy

https://chatgpt.com/share/63fc4ae8-ddc2-4850-b88a-4113ff5a235b

A2.2 Compensation involved

https://chatgpt.com/share/1d3565d8-55d6-4b0a-84e7-46daca3be2da

A2.3 Roles involved

https://chatgpt.com/share/1a19f694-fb46-4638-9c7a-3a58b8834c67

A2.4 The use of a specific action and prohibition

https://chatgpt.com/share/441ddd89-e873-43c9-a945-f2ca9c5f8991

A2.5 Basic constraints

https://chatgpt.com/share/b6044ab6-0331-4ff2-a3f5-1fc8dcad027c

A2.6 Mismatch example

https://chatgpt.com/share/8365e7e5-ab7c-49a5-9d38-6d595485e942

A2.7 Combination of former scenarios

https://chatgpt.com/share/432fc211-0526-4115-8013-df8281d19bbf

A2.8 A real life FDT use case

https://chatgpt.com/share/9fb6269c-eae6-4552-bd8f-6e5a734b2749

A2.9 A non-specified asset in the request

https://chatgpt.com/share/4107fa02-5366-4b25-beb1-1faada4aaf10

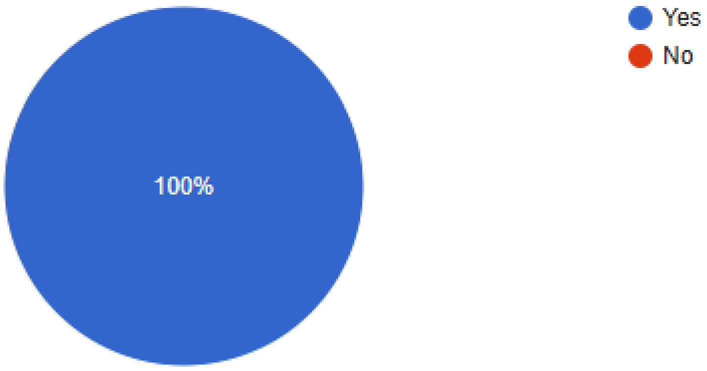
A2.10 Agreement based on the most common access conditions in health data.

https://chatgpt.com/share/219296a5-1b86-4733-9a7a-3d0b1e0e41d2

C Survey Response

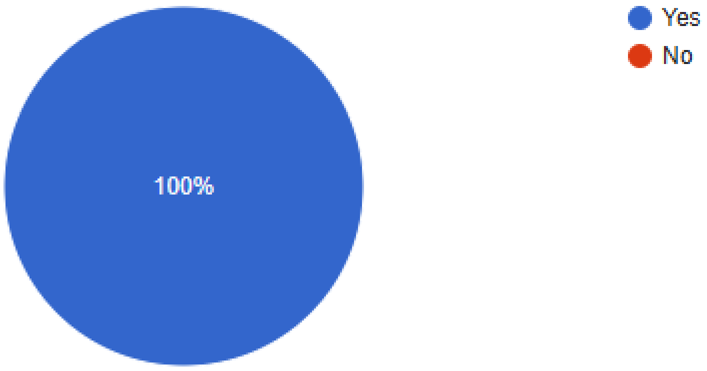
Are you able to interpret Resource Descriptive Frameworks (RDF)?

1 answer



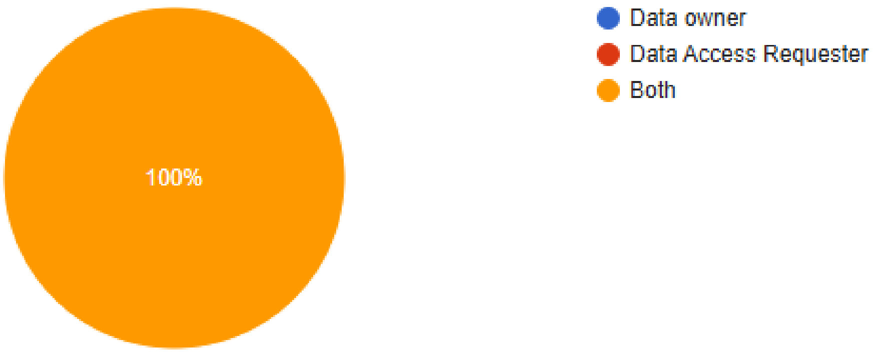
Have you ever had to deal with data access conditions?

1 answer



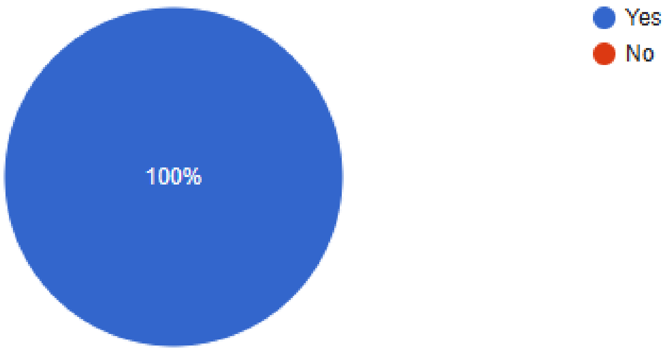
What was your role when dealing with data access conditions?

1 answer



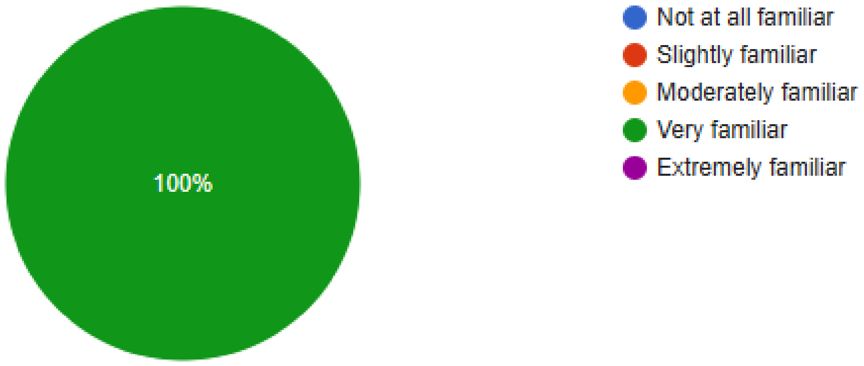
Are you involved in the collaboration between the University of Twente and farmers in a FAIR Data Train project?

1 answer



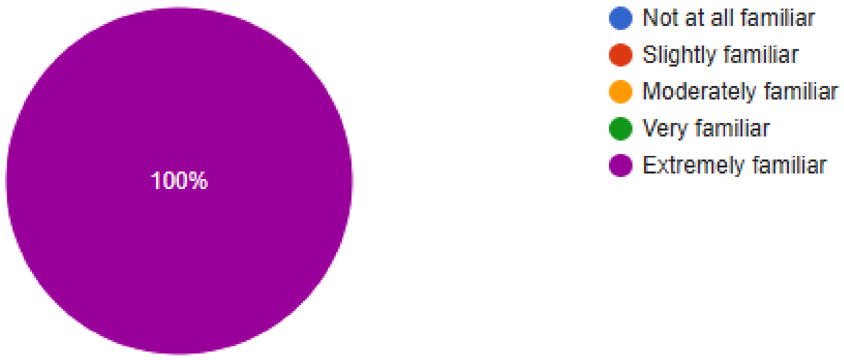
How would you rate your familiarity with linked data or Resource Description Framework (RDF)?

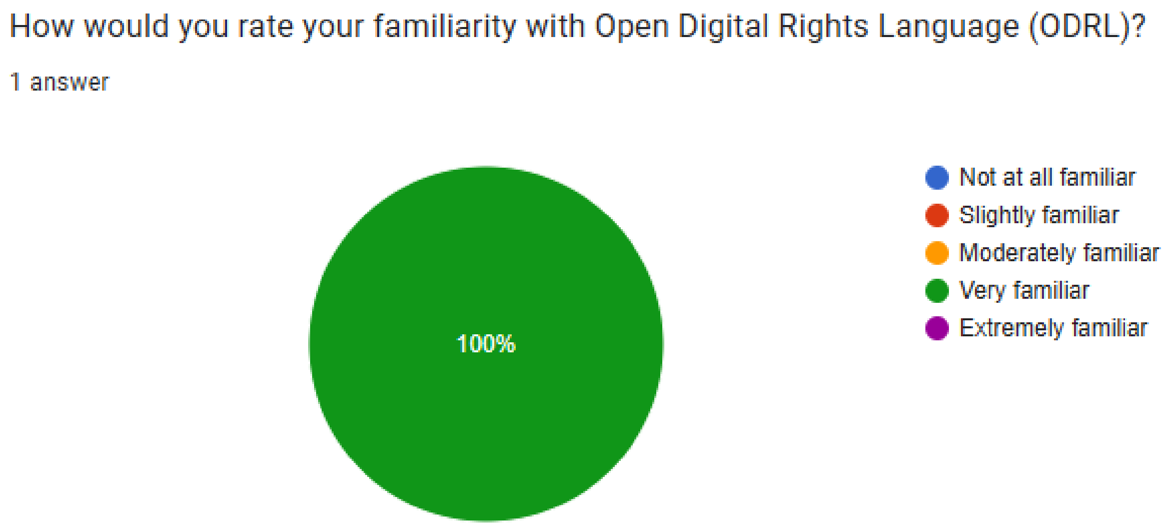
1 answer



How would you rate your familiarity with the FAIR Data Train/Personal Health Train?

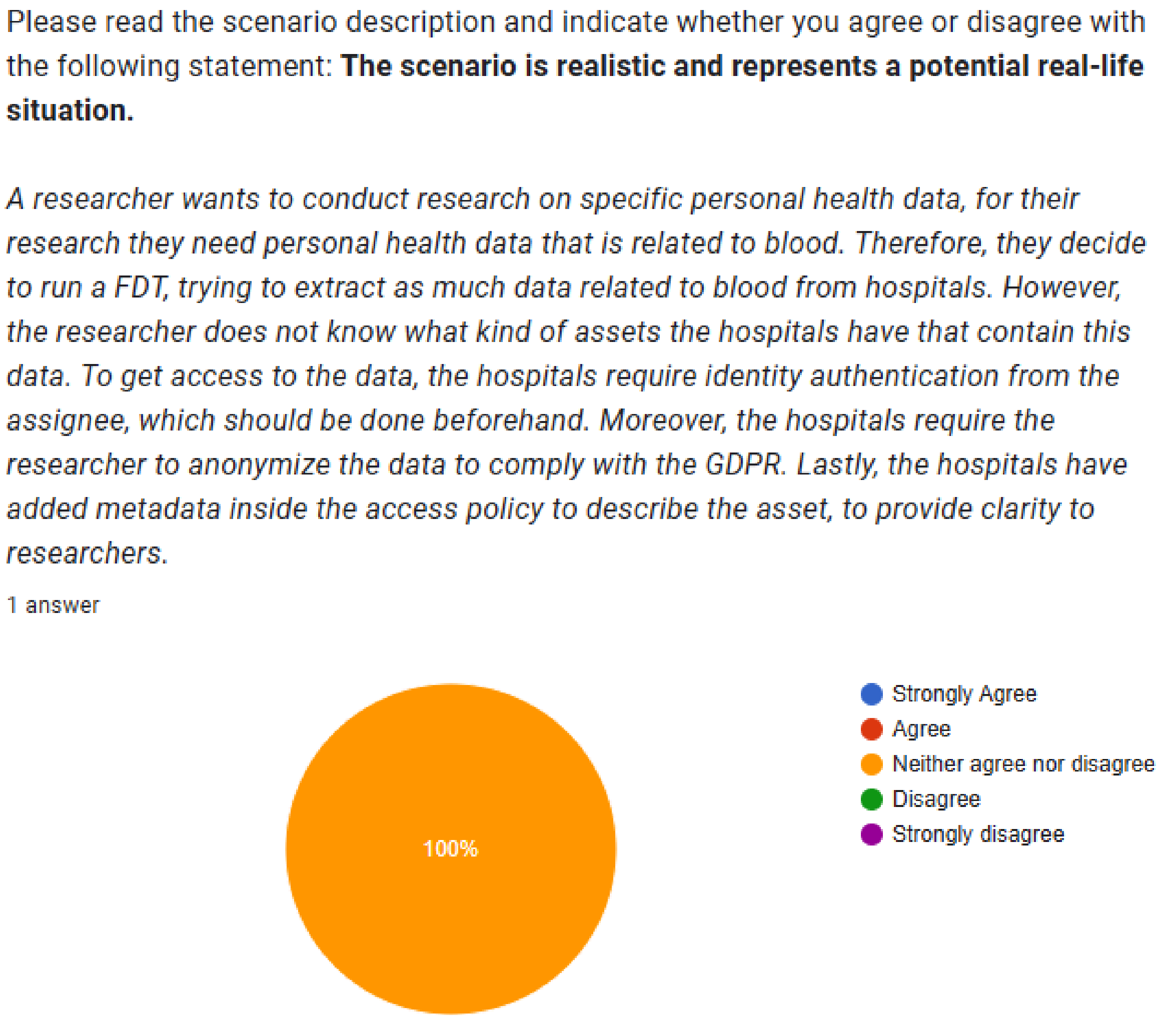
1 answer





How would you rate your familiarity with Open Digital Rights Language (ODRL)?

1 answer



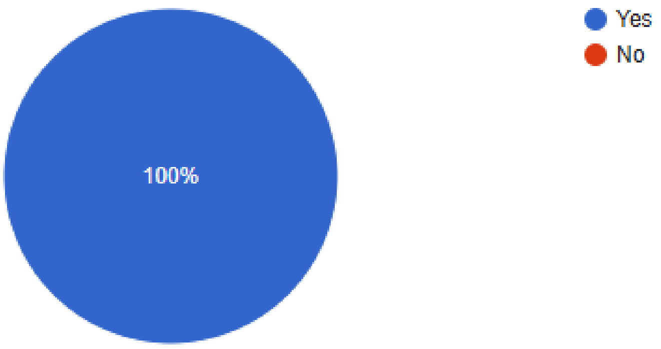
Please read the scenario description and indicate whether you agree or disagree with the following statement: **The scenario is realistic and represents a potential real-life situation.**

A researcher wants to conduct research on specific personal health data, for their research they need personal health data that is related to blood. Therefore, they decide to run a FDT, trying to extract as much data related to blood from hospitals. However, the researcher does not know what kind of assets the hospitals have that contain this data. To get access to the data, the hospitals require identity authentication from the assignee, which should be done beforehand. Moreover, the hospitals require the researcher to anonymize the data to comply with the GDPR. Lastly, the hospitals have added metadata inside the access policy to describe the asset, to provide clarity to researchers.

1 answer

Have you ever experienced a situation similar to the scenario described above where a researcher does not know what data they should access?

1 answer

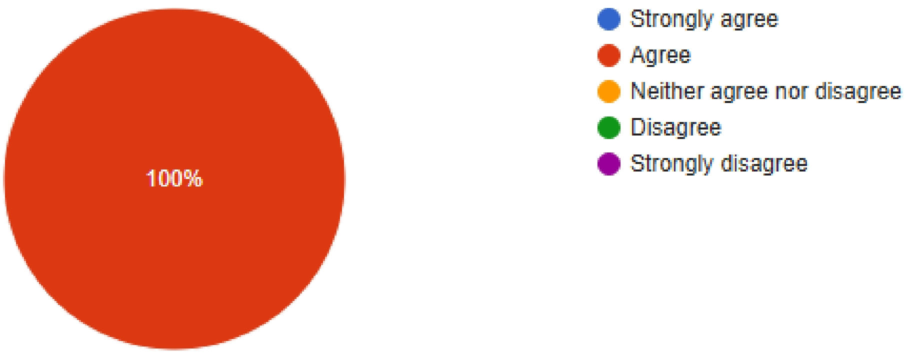


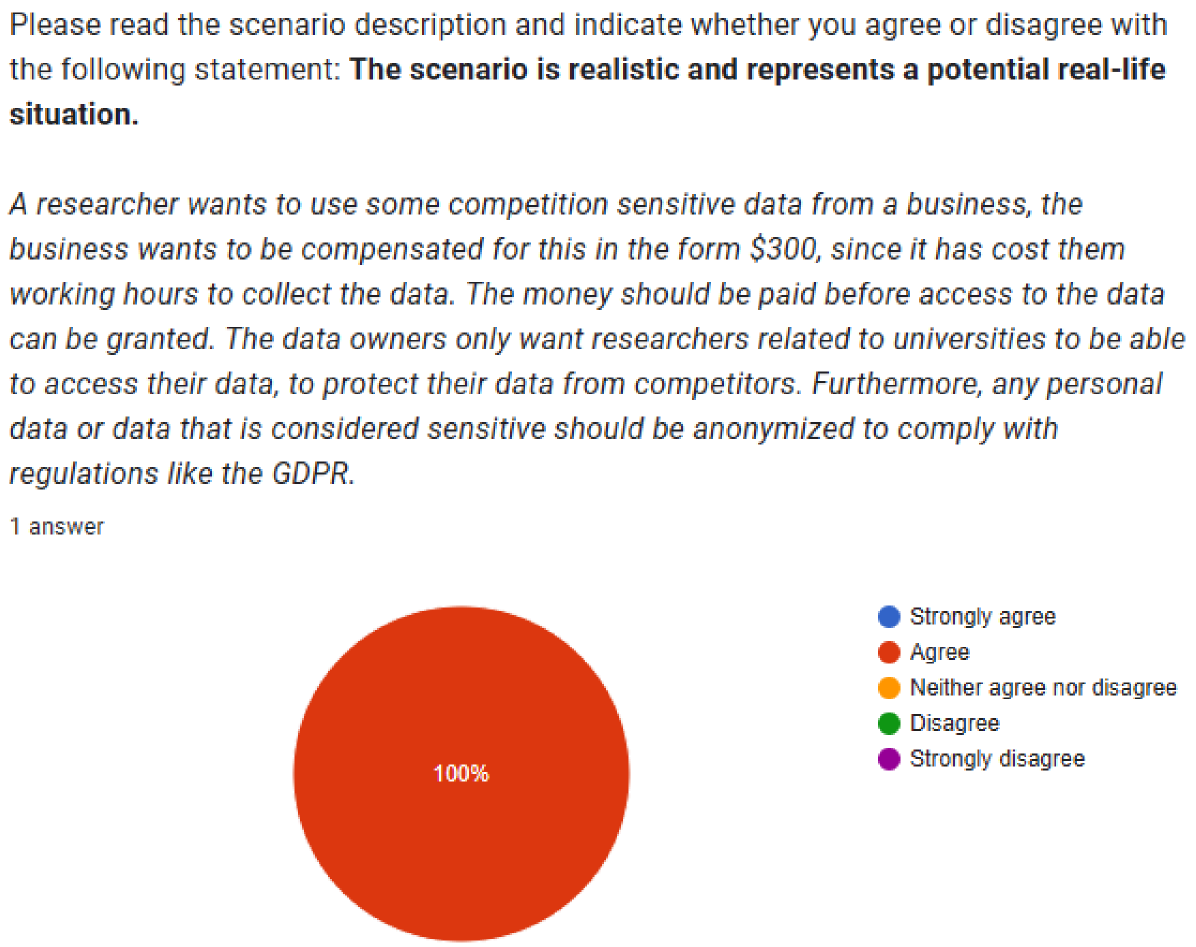
Please analyze the following RDF and indicate whether you agree or disagree with the following statement: **The RDF represents an agreement based on the scenario described in the previous question.**

Description: The agreement stipulates that party B is permitted to use a specific asset containing personal health data under the following conditions:
The assignee (researcher/ the party that requests access) must include a media file that authenticates their identity.They must adhere to the constraint that identity authentication must be done before access is granted. Additionally, they have a duty to anonymize the data.

RDF:

1 answer

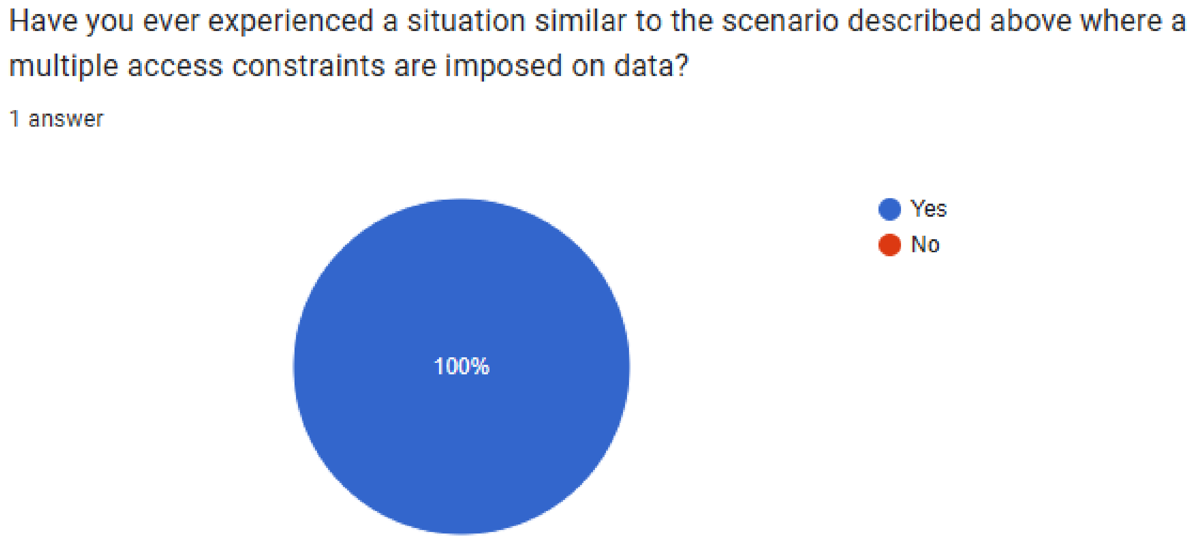




Please read the scenario description and indicate whether you agree or disagree with the following statement: **The scenario is realistic and represents a potential real-life situation.**

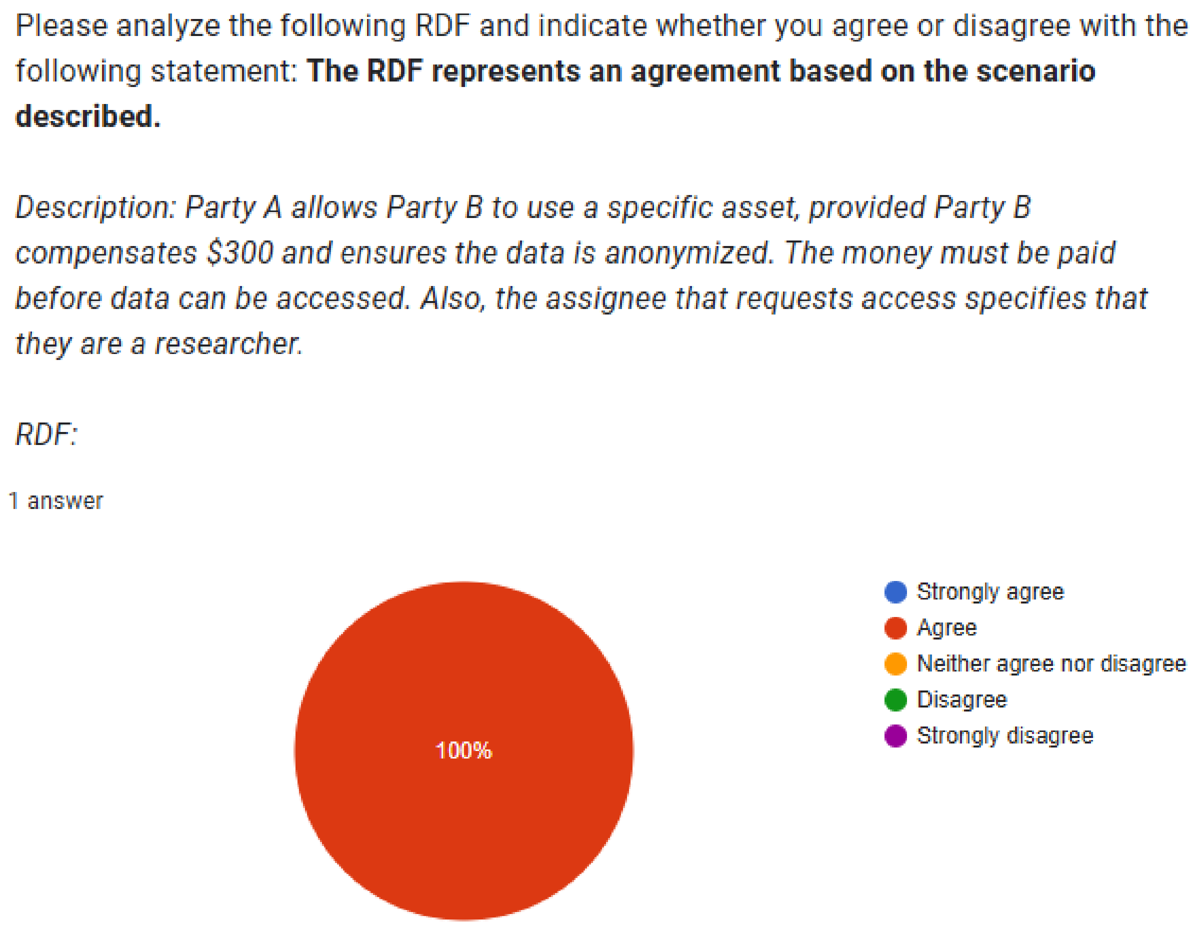
A researcher wants to use some competition sensitive data from a business, the business wants to be compensated for this in the form $300, since it has cost them working hours to collect the data. The money should be paid before access to the data can be granted. The data owners only want researchers related to universities to be able to access their data, to protect their data from competitors. Furthermore, any personal data or data that is considered sensitive should be anonymized to comply with regulations like the GDPR.

1 answer



Have you ever experienced a situation similar to the scenario described above where a multiple access constraints are imposed on data?

1 answer

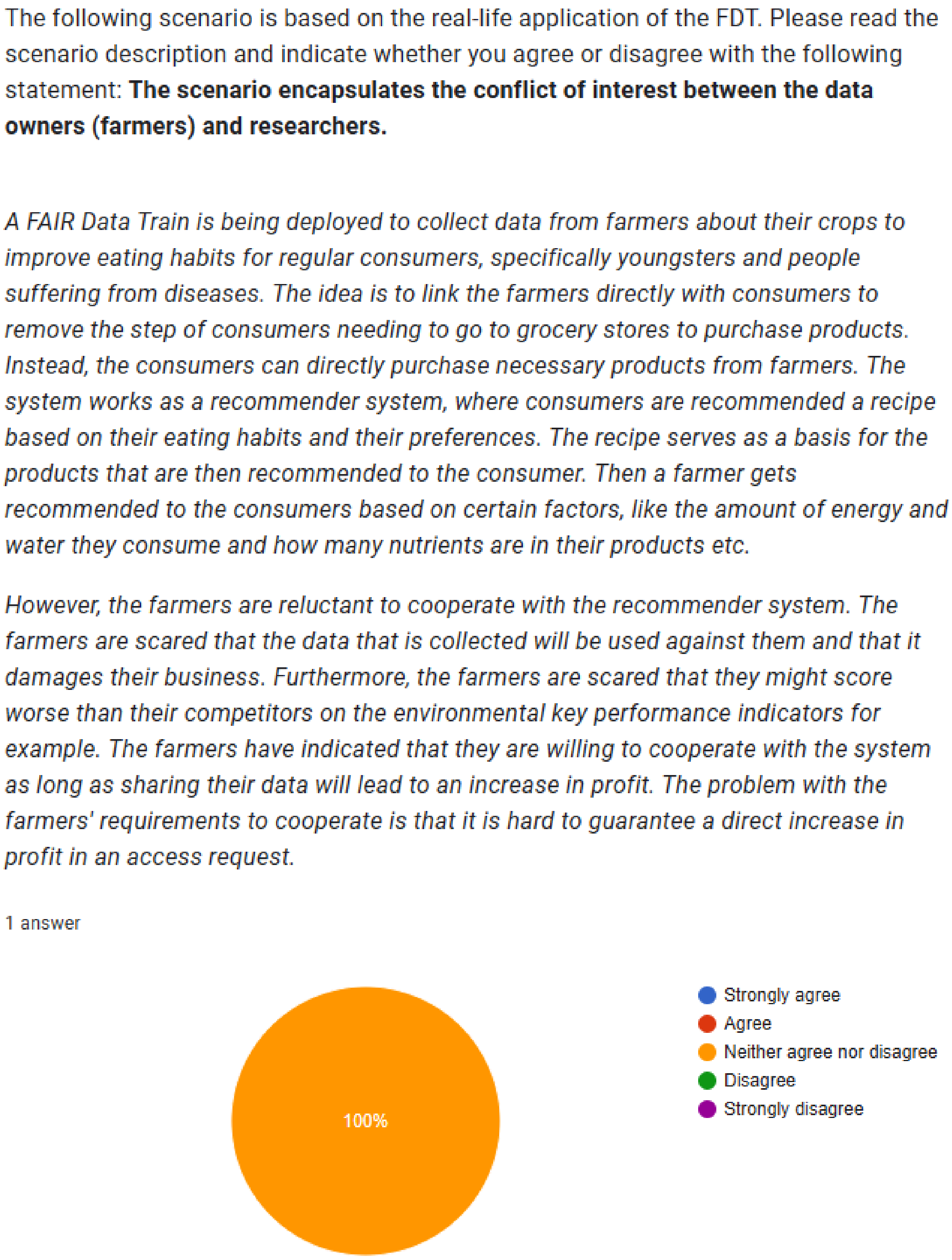


Please analyze the following RDF and indicate whether you agree or disagree with the following statement: **The RDF represents an agreement based on the scenario described.**

Description: Party A allows Party B to use a specific asset, provided Party B compensates $300 and ensures the data is anonymized. The money must be paid before data can be accessed. Also, the assignee that requests access specifies that they are a researcher.

RDF:

1 answer



The following scenario is based on the real-life application of the FDT. Please read the scenario description and indicate whether you agree or disagree with the following statement: **The scenario encapsulates the conflict of interest between the data owners (farmers) and researchers.**

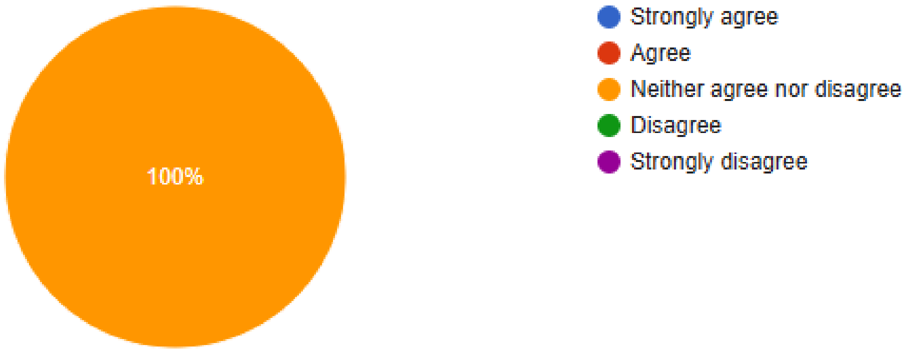
A FAIR Data Train is being deployed to collect data from farmers about their crops to improve eating habits for regular consumers, specifically youngsters and people suffering from diseases. The idea is to link the farmers directly with consumers to remove the step of consumers needing to go to grocery stores to purchase products. Instead, the consumers can directly purchase necessary products from farmers. The system works as a recommender system, where consumers are recommended a recipe based on their eating habits and their preferences. The recipe serves as a basis for the products that are then recommended to the consumer. Then a farmer gets recommended to the consumers based on certain factors, like the amount of energy and water they consume and how many nutrients are in their products etc.

However, the farmers are reluctant to cooperate with the recommender system. The farmers are scared that the data that is collected will be used against them and that it damages their business. Furthermore, the farmers are scared that they might score worse than their competitors on the environmental key performance indicators for example. The farmers have indicated that they are willing to cooperate with the system as long as sharing their data will lead to an increase in profit. The problem with the farmers' requirements to cooperate is that it is hard to guarantee a direct increase in profit in an access request.

1 answer

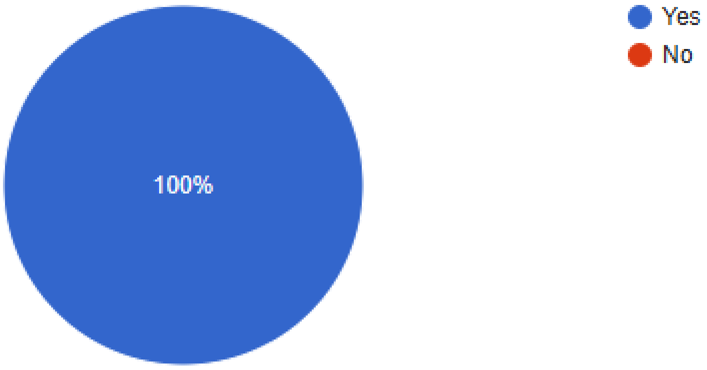
Please, based on the scenario described below, indicate whether you agree or disagree with the following statement: **The scenario accurately describes why the data owners (farmers) are reluctant to share their data.**

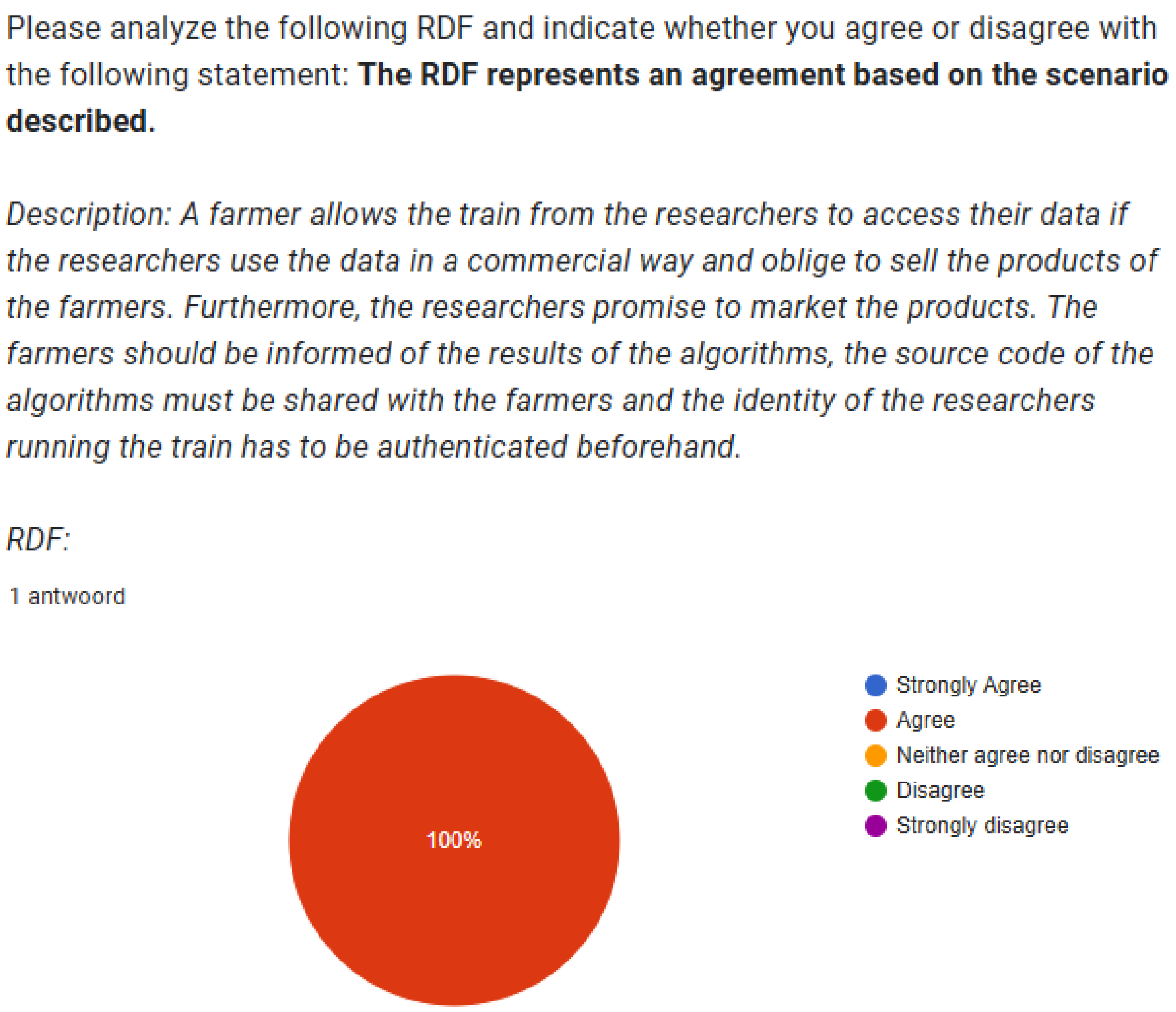
1 antwoord



Have you ever experienced a situation similar to the scenario described above where data owners involved are reluctant to share data?

1 antwoord



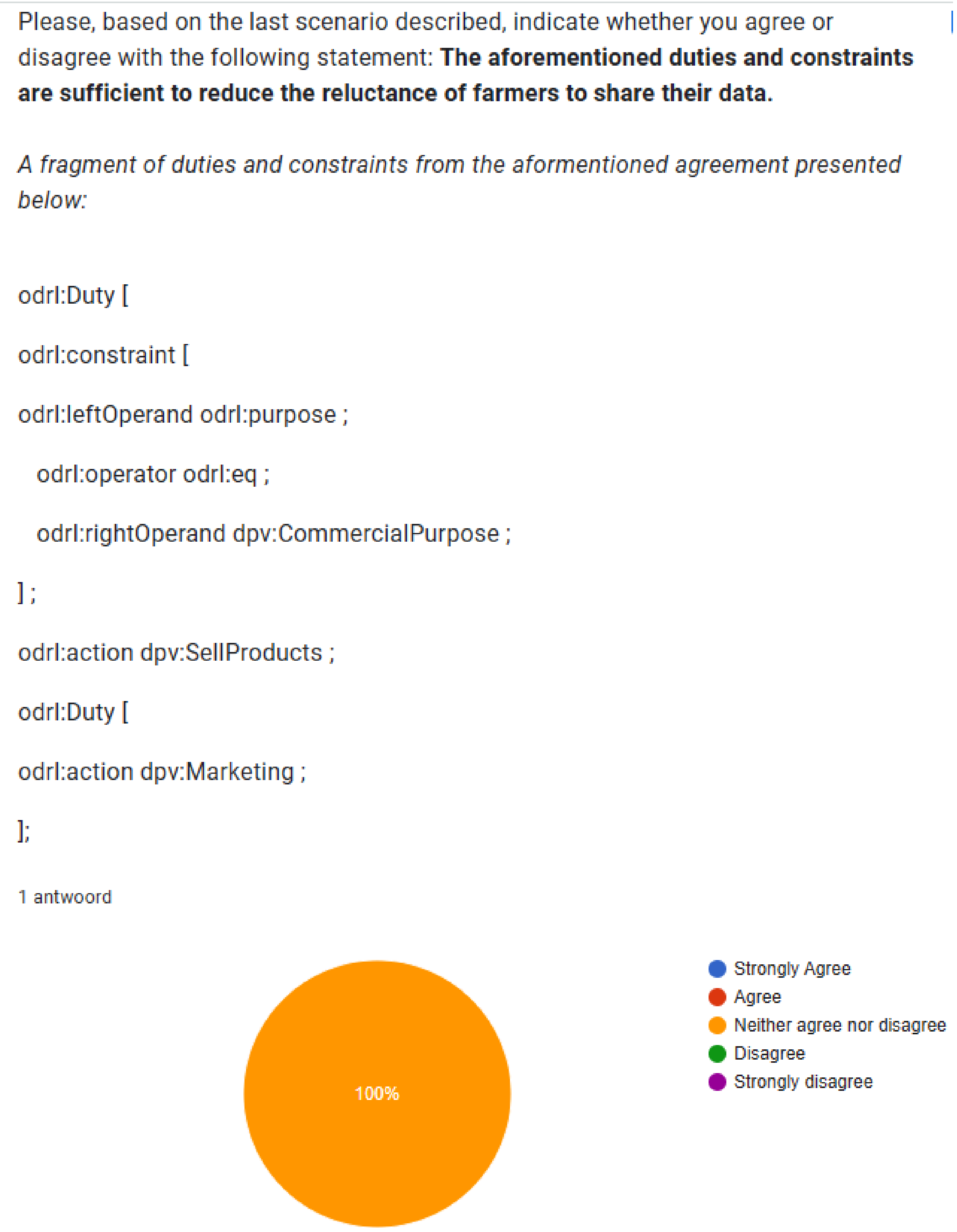


Please analyze the following RDF and indicate whether you agree or disagree with the following statement: **The RDF represents an agreement based on the scenario described.**

Description: A farmer allows the train from the researchers to access their data if the researchers use the data in a commercial way and oblige to sell the products of the farmers. Furthermore, the researchers promise to market the products. The farmers should be informed of the results of the algorithms, the source code of the algorithms must be shared with the farmers and the identity of the researchers running the train has to be authenticated beforehand.

RDF:

1 antwoord



Please, based on the last scenario described, indicate whether you agree or disagree with the following statement: **The aforementioned duties and constraints are sufficient to reduce the reluctance of farmers to share their data.**

A fragment of duties and constraints from the aformentioned agreement presented below:

```
odrl:Duty [

odrl:constraint [

odrl:leftOperand odrl:purpose ;

  odrl:operator odrl:eq ;

  odrl:rightOperand dpv:CommercialPurpose ;

] ;

odrl:action dpv:SellProducts ;

odrl:Duty [

odrl:action dpv:Marketing ;

];
```

1 antwoord