DataConsensus - A Collaborative and Interoperable Tool for Data Cooperatives

Bríd O'Donnell, BAI

A Dissertation

Presented to the University of Dublin, Trinity College

in partial fulfilment of the requirements for the degree of

Master of Science in Computer Science (Intelligent Systems)

Supervisor: David Lewis

August 2023

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

Bríd O'Donnell

August 18, 2023

DataConsensus - A Collaborative and Interoperable Tool for Data Cooperatives

Bríd O'Donnell, Master of Science in Computer Science

University of Dublin, Trinity College, 2023

Supervisor: David Lewis

Despite the wealth of data within the world, health researchers struggle with high regulatory and technical burdens to gather data necessary for their work. Enabling more data sharing while preserving rights and privacy is the key objective of the European Strategy for Data and the focus of this research. Specifically this work tackles the question "How to enable data cooperatives to collaboratively deliberate and decide on interoperable data requests from third parties?" by proposing the DataConsensus application, a tool built on Solid to manage data requests for a group of Nightscout users. While tailored for a specific community, the project's contributions can be adapted to other scenarios due to its interoperability emphasis. This work proposes the ODRL Profile for Collective Policies (OCP), to define and track policies within the DataConsensus system using existing vocabularies including the Digital Rights Language. Along with the use of the Solid specification, this ensures the interoperability and transparency of the application. Additionally, the application promotes collaborative decision-making through its proposed decision process, allowing members to propose counter offers to third parties. Furthermore, this work is evaluated against its data policy checklist and community related guidelines, which can be utilised against future research. In conclusion, the project has made significant contributions to research on data cooperatives.

Acknowledgments

I would like to take this opportunity to thank my supervisor, Professor Dave Lewis, for his support and guidance throughout this project. I am deeply grateful for his enthusiasm and his insightful pointers for my project. I would also like to thank Beatriz Esteves, an Early Stage Researcher (ESR) working on PROTECT (Protecting Personal Data Amidst Big Data Innovation), who gave me early guidance in this project.

Furthermore, I'd like to thank my mum and dad for their support throughout this dissertation. I would also like to acknowledge my older sister, who has had firsthand experience with the burdensome process of collecting health data for research. She motivated me to pursue this project and I am very thankful she did. This dissertation was a challenge but a worthwhile one and I learnt a great deal during the process.

Bríd O'Donnell

University of Dublin, Trinity College
August 2023

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Despite predictions that the global data volume will grow from 33 zettabytes to 175 zettabytes [10], much of the world's data is restricted to silos and its full potential is never used. Major obstacles to free flowing data include the lack of infrastructure and interoperability, and governance uncertainty. For researchers who wish to utilise health data, these obstacles compound further with higher burdens [19]. This project is addressing this problem through data cooperatives. Specifically, this project proposes DataConsensus, a tool that enables groups of individuals to pool their data, deliberate and vote on data requests, and grant access to the pooled data to the pooled researchers. The aim of the DataConsensus application is to bridge the gap between the rights of an individual to control their data and the benefits of sharing data for research.

1.1 Motivation

Data Governance has been a key area of regulation within the EU since the implementation of GDPR in 2018. While most of these efforts have been concentrated in enhancing individuals' right over their personal data, the European Strategy for Data, revealed in 2020, demonstrates that the European Commission also hopes to build a functioning market and ecosystem for data sharing while protecting individual's rights [10]. One of the motivations behind this strategy is the chilling effect GDPR and its corresponding national regulations had on legitimate data sharing for research or altruistic purposes. Particularly, for areas with sensitive personal data such as healthcare, gaining access to large amounts of data ethically is a significant obstacle for researchers pursuing worthwhile research [19]. Therefore, balancing individual's rights and the benefits of data-sharing is a major challenge for the EU and a core consideration for this project.

The changing regulatory landscape has opened the door to new alternative data gov-

ernance models, challenging the status quo of individual due diligence and obscure data marketplaces. One such idea is a data cooperative - a group of data subjects pooling their data and decision-making in order to benefit from shared due diligence and increase their leverage towards parties desiring their data [82]. Given this collaborative nature, the challenges of deliberation will need to be overcome to ensure optimal trust in the governance of the data cooperative. This will be a key focus on the project.

Finally, there is a growing body of work into standardising data requests and consent agreements into machine readable formats. The desired benefits of interoperability is that it will allow for greater automation while still preserving privacy in these interactions. One approach has been utilising the Online Digital Rights Language (ODRL) and Digital Privacy Vocabulary (DPV), as proposed by [33] with their ODRL Profile for Access Control. This technical implementation will serve as the foundation of this project.

1.2 Research Question

The research question this research aims to answer is formulated as follows:

"How to enable data cooperatives to collaboratively deliberate and decide on interoperable data requests from third parties?"

1.3 Research Approach

To answer the above research question, the following approach was taken to complete this project:

1. Identify the legal requirements for data sharing platforms;
2. Determine best practices in data cooperative tools and deliberation processes;
3. Formulate a realistic setting and related user stories for a data cooperative;
4. Create an ODRL profile for data cooperatives enabling traceable and collaborative decisions making between members and third parties;
5. Design and develop a data cooperative application, built using SOLID and the proposed ODRL profile;

1.4 Research Approach

This research contributes to the Solid community, the ODRL community and Data Cooperative literature. The specific contributions of this research are the following:

1. A BPMN decision process diagram for data cooperatives, which proposes a multiple stage collaboration between members and third parties. This contribution will be introduced in Chapter 3.

2. The ODRL profile for collective policies (OCP) provides an interoperable method for the collaborative decision process proposed in this work to be recorded and tracked. This was built upon previous work from [33] and [91]. This contribution will be introduced in Chapter 3.

3. The DataConsensus application, an application built on solid, that implements the proposed decision process and ODRL profile in a practical example. Chapter 4 describes the implementation of this application.

4. A data policy checklist, covering the legal requirements and best practices for obtaining consent, specifically for data commons sharing health data with researchers within Ireland. This was used to evaluate the coverage of the OCP and can serve as an evaluation tool in further research into health related data sharing initiatives. This contribution is presented in Chapter 5.

5. A list of guidelines for designing and building a fair and sustainable data commons. These were extrapolated from the existing literature and can serve as both guidance and an evaluation tool for future work into data commons. This contribution is presented in Chapter 5.

1.5 Non-Goals

In this section, the non-goals of this project will be outlined. The application developed is to serve as a proof-of-concept system for demonstrating capabilities, rather than a production-ready system. While high standards of coding practices were observed for the DataConsensus application, scalability and performance is not the focus of the application.

Furthermore, this research does not aim to evaluate the user experience of the application as no patient groups were identified and educated to test the application. While I have studied the technical aspects of Solid to the extent required to build the DataConsensus application, this report only provides a basic understanding for solid and is not a complete guide to Solid.

Finally, the application designed in this project is not designed for individuals who lack the capacity to grant consent, such as people under the age of 18 years old. Furthermore, this project will not incorporate revenue sources or monetary exchanges in its designs for a data cooperative.

1.6 Report Structure

Throughout this document, the words this research, this work and this project are used interchangeably to refer to the work done towards the completion of this dissertation, unless specified otherwise. The words the application, and DataConsensus are used interchangeably to refer to the application developed as a part of this dissertation.

This report is organised into the following:

Chapter 2 explores the three research areas associated with this project: the relevant legislation, data governance models and data cooperatives, and SOLID and existing ontologies to support this work.

Chapter 3 discusses the design process and choices for the application, as well as the explanation for the deliberation and voting process developed in this project.

Chapter 4 explains the technical implementations of the application, including architecture, set-up, as well as a detailed description of the ODRL profiles.

Chapter 5 reviews the application developed by this research, evaluating it against the goals achieved, legal compliance and existing guidelines. This chapter also presents a security and privacy evaluation of the application.

Chapter 6 summarises this research and presents the conclusions drawn from it. The chapter concludes with the future work for improving the application and areas for future research in the relevant literature.

Furthermore, Appendix A describes the use of AI in this project, Appendix B presents three examples of OCP policies in turtle format and Appendix C covers instructions for setting up the application and other development guides.

Chapter 2

Related Work

In this Chapter, the related work and key background to this project will be discussed. Section 2.1 will describe why the research areas were chosen and the strategies used to investigate the literature. Section 2.2 will describe the legislation and regulation that govern data sharing. Section 2.3 explores the research area of data governance and data commons. Section 2.4 examines the background of Solid and privacy related policy languages and vocabularies. Finally, Section 2.5 concludes with a summary of the key insights gathered from the related works.

2.1 Literature Review Strategy

Figure 2.1 illustrates the three research areas that intersected in this project. These areas were chosen as the legislation reflects the legal requirements for data sharing initiatives and how it will change in the future. The literature on data governance and data commons provides guidance on specific flow and functionalities an application such as DataConsensus should have. Finally, SOLID and related linked data technologies are relevant because they provide the practical tools and building blocks for implementing the insights discovered from the first two areas.

Multiple strategies were employed for the literature review in this research such as snowballing [113]. The search began with a number of key documents and the other relevant documents were discovered through the bibliographies of those initial documents. After an initial snowballing, a list of key search terms were identified that encapsulated the key concept and then searched in the online database Google Scholar. This combined strategy ensured that the literature review was both exhaustive and relevant.

When covering the legislation, the official European Commission and adjunct sites were key sources for gathering information. Additionally, a number of news articles were

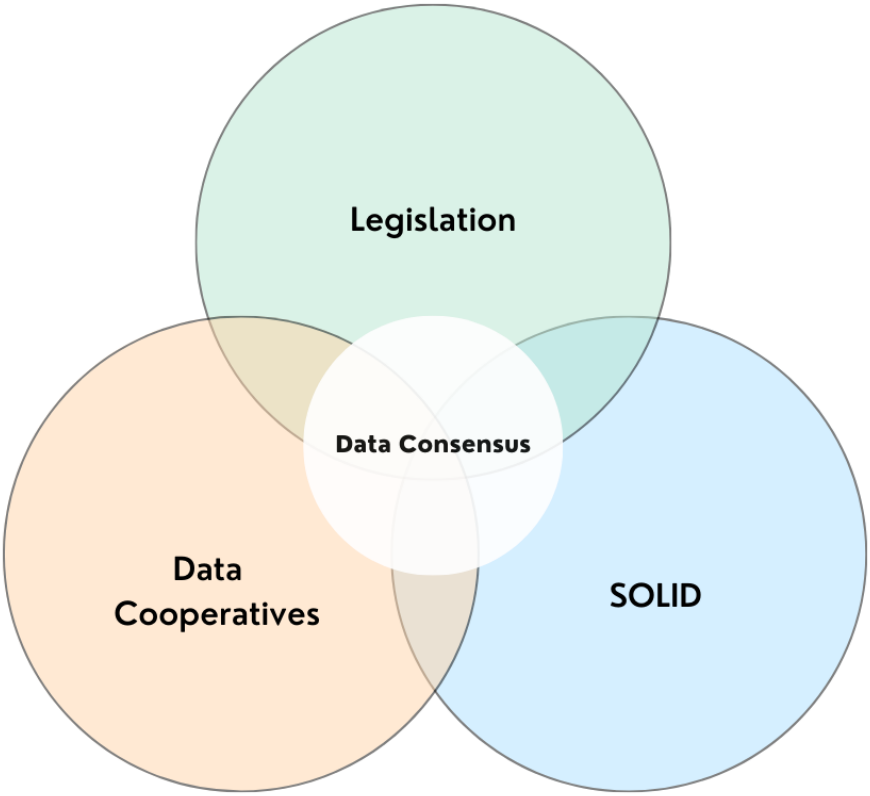


Figure 2.1: The intersecting research area for the Data Consensus project

sourced to better understand the status and challenges to the legislation. Furthermore, the key sources of information regarding Solid were the official Solid website and Inrupt Inc, the dominant library provider for solid. The existing data sharing initiatives were initially identified from the literature and then further investigated through their websites and related media.

2.2 Legislation and Regulation

In this section, we will explore the legislation and regulations relevant to this project. This will include European legislation and national Irish legislation. Specifically, the legislations and regulations examined are the General Data Protection Regulation (GDPR) [39] and Ireland's Health Research Regulation, as well as the EU's Data Act [36] and Data Governance Act (DGA) [40]. We will also discuss the European Strategy For Data. This section will examine the actual text of the legislation, along with official guidance and other literature critiquing the laws and illustrating the real-world impact of the laws. The objectives of reviewing this content is twofold:

1. The legal requirements outlined in the legislation and regulations discussed here will serve as an evaluation tool in chapter 5.

2. The legislation and regulations influence the landscape of data governance and data spaces.

2.2.1 GDPR

The General Data Protection Regulation (GDPR) came into effect across the European Union in 2018, with the aim to protect EU citizen's privacy and data. This law is relevant to this project as it sets out specific requirements to comply with when dealing with personal data. The legislation in fact recognises special categories of personal data, such as data concerning health [25], which falls into this project's scope. GDPR also specifies six different legal bases for the processing of personal data [19]. The DataConsensus application utilises one of these legal bases - consent, which will be discussed in more depth below.

Furthermore, GDPR defines a number of roles relevant to the processing of personal data and these roles are relevant when defining the ODRL policies of this project. Those roles are:

- A data subject is an individual whose personal data is being processed. GDPR has granted a number of rights to data subjects, including the right to erase of personal data (Article 17) and the right to object to processing personal data (Article 21) [19].

- A data controller is an individual or legal person(s) such as a company, department, or organisation, which under Article 4 of GDPR "determines the purposes and means of the processing of personal data" [39].

- There is also the term "joint controller", which allows for multiple data controllers, who all determine the purpose and means of data processing but may not imply equal responsibilities [14].

There are other roles outlined too. A recipient is an entity to whom personal data is disclosed, while a third party is an entity "other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data" according to Article 4 of GDPR [39]. Recipient and third party will be used later in this report while discussing the DataConsensus but in those cases, they are not representing their legal terms. A data processor is also defined

in the legislation, however they aren't particularly relevant in the current state of the DataConsensus application.

The core rules regarding personal data processing are presented in Article 5 of GDPR [39]. Additionally, in cases where the personal data processing is being performed based on the legal basis of consent (as is the case in the application), the data subject shall have the right to withdraw consent at any time and withdrawing should be as easy as it is to grant consent. These rules dictate key requirements for the DataConsensus application.

Processing of special categories of personal data, including health data, can occur when explicit consent has been given by the data subject according to Article 9(2)(a) of GDPR [39]. GDPR does not define explicit consent but defines consent in Article 4 as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by clear affirmative action, signifies agreement to the processing of personal data relating to him or her" [39]. Consent is only one legal basis by which to pursue health research as researchers can also use the legal basis of scientific research in the public interest [39]. However, explicit consent will remain relevant in Ireland and further definitions of explicit consent along with other outstanding questions about the practical implementation of GDPR in national legislation and regulations, such as Ireland's Health Research Regulation.

2.2.2 Irish Context

Ireland's Health Research Regulations (HRR) was introduced in 2018 following GDPR and regulates the processing of personal data for health research in Ireland. It defines health research, which includes experimental, translational and clinical research, population health research, public health and social care research, as well as research into treatment strategies, medical device or product development [16]. Another key point of the HRR is making explicit one of the mandatory "suitable and specific measures" that must be in place for the processing of data for health research purposes, barring certain exceptions, differentiating it from GDPR and other EU countries.

The introduction of HRR was welcomed by the Irish medical field but they also identified significant challenges in the law, specifically regarding explicit consent. According to the Department of Health guidance, explicit consent is consent that has been documented and such documentation is shared with the individual. A later amendment to the law specifies that the consent must be obtained "in accordance with international best practice on the ethical conduct of health research (which includes informed consent, transparency and independent ethical oversight)" [44]. As such, guidance from Health Research Board, states that an explicit consent statement should refer to:

- "the particular data set that is to be processed,
- the precise purpose of processing (including any automated decision-making),
- should identify any risks and/or implications that might arise for the data subject as a result of the data processing, and
- should provide any other relevant and specific information that might influence the decision of a data subject to give or not give their consent." [16]

Along with the other basic requirements required in accordance with GDPR's definition of consent, as outlined in the previous section. It should be noted that the HRR does allow for individuals to give "broad consent", essentially granting explicit consent when the research area is only generally defined, however, this is not encouraged ([16]). The mention of informed consent is another definition to explore. Informed consent was a concept used before the existence of GDPR and was a requirement for conducting research in most circumstances. With the inclusion of both explicit and informed consent, the HRR essentially requires double consent from patients, a confusing and burdensome requirement for researchers. Explicit and informed consent and the guidelines for how to obtain them will serve as an evaluation tool for this application.

There was also significant confusion over understanding the HRRs and leading to "huge site to site variation" in its implementation. There was a negative impact on clinical trials, biobanks and training programmes. As the HRR differentiated Ireland from other EU countries, even international trials between EU countries were stalled and burdened. The general view among individuals involved in health research was that the HRR has placed "a significant extra burden of work on Ireland's clinical researchers", and has had a negative effect on Irish research [19] [65].

Several of the identified concerns have been mitigated through revisions to the HRR, implemented in 2021 [44]. However, the complexities presented by the Health Research Regulations (HRR) underscore the necessity for reform and standardisation in the process of obtaining consent from data subjects. Such reform and standardisation should alleviate the pressures on researchers, thereby facilitating faster and more research. More standardisation would also enhance the patients' understanding of the consent they provide. Additionally, it would likely foster more international cooperation in health research with other EU countries.

2.2.3 European Strategy For Data

The European Strategy for Data demonstrates the EU's recognition of data as a key resource in today's society. The strategy hopes to bolster Europe's competitiveness as

a player in the data economy, while still ensuring consumer protections, security and a strict legal framework is in place, a difficult balance to strike. The goals related to increasing Europe's competitiveness include allowing for more data to flow across EU borders, investing in security, cloud and data processing infrastructure and creating more data sharing initiatives and frameworks [34].

A new data sharing concept introduced by the EU Data Strategy was the Common European Data Spaces. When the Strategy was revealed in February 2020, the creation of 10 data spaces in key strategic fields, including energy, agriculture and as mentioned above, health [34]. According to the Commission Staff Working Document on Common European Data Spaces, "Common European data spaces bring together relevant data infrastructures and governance frameworks in order to facilitate data pooling and sharing" [37].

The European Health Data Space (EHDS) is one of these Common European Data Spaces proposed by the Data strategy and aims to standardise the rules, standards and infrastructure within the healthcare space. One of the key objectives of the EHDS is to allow Researchers, policymakers and companies to see the pseudonymised and anonymised data from patient's electronic health records if they receive a permit from a data access body [75].

However the exact governance and scale of this access is unclear and the timeline for implementing the EHDS will be longer than initially planned [84]. This is due to concerns surrounding privacy, as well as challenges in the digitalisation and harmonisation of patient information systems, which vary significantly at both the state and regional levels.

Regardless, the EU Data Strategy indicates the EU's willingness to develop data sharing initiatives and the EHDS represents an ambitious goal to make sharing health data easy and acceptable. Two more key pillars to the EU Data Strategy is the Data Act and the Data Governance Act (DGA) which is discussed further in the sections below [35]. Furthermore, the EU Data Strategy and its legislation make a wider landscape of technology-focused legislation including Digital Markets Act, the Digital Service Act and the AI Act, but they are beyond this report's purview.

2.2.4 Data Act

The main goal of the Data Act is making more data available for use. From a high level perspective, the Data Act is tackling who controls the data generated by connected devices and the law ensures that individuals have the right to access the data they generate on connected devices, free of charge and in real-time [48]. The right to access the data

you generate is relevant to this project as it means patients should always be able to easily access data belonging to their medical wearables and then share that data with third parties or in the case of the DataConsensus application, pool the data with other wearable users.

Another objective of the Data Act is providing the legal basis for developing interoperability standards for data use and sharing [36]. While one motivation stated for this is the increased competition in the cloud market, it will also remove barriers for Common European Data Spaces and other kinds of data sharing. Commentary on the Data Act does suggest that any interoperability framework "needs to produce the conditions for creating a semi-common perspective, cooperative data government and public data trust" and warns of fears that the development of any standards could be easily overtaken by established private interests [94]. Regardless, interoperability is a key consideration for this project, which is why linked data and RDF is a core component of the DataConsensus application.

Along with interoperability standards for data sharing, the Act also addresses smart contracts, defined broadly as "a computer program stored in an electronic ledger system wherein the outcome of the execution of the program is recorded on the electronic ledger". The Act specifically encourages the standardisation of smart contracts for dating sharing applications. In the public online consultation on the Data Act, "79% of respondents considered that smart contracts could be an effective tool to technically implement data access and use in the context of co-generated Internet of Things data" [36]. This project doesn't involve smart contracts but it does propose an alternative to smart contracts in the form of its ODRL profile and the DataConsensus application. Much like smart contracts, the standardisation of ODRL profiles and Linked Data access control methods should be promoted [71].

Another proposal from the Data Act relevant to this project is a set of requirements for Data Spaces to facilitate interoperability of data, data sharing mechanisms and services. These requirements will be incorporated into the evaluation measures in Chapter 5. Additionally, the Data Act defines some roles that will be used in this report. A data holder is defined as a person or organisation who has the right or obligation to share specific data, while the data recipient is a person or organisation who receives data from a data holder [23].

2.2.5 Data Governance Act

There are two key concepts introduced in the Data Governance Act (DGA) relevant to this project: data intermediation services and data altruism organisations. The term Data

Intermediary is defined by the internet policy review as "a mediator between those who wish to make their data available, and those who seek to leverage that data" [58]. The DGA narrows this definition to "a service, which aims to establish commercial relationships for the purpose of data sharing between an undetermined number of data subjects and data holders, on the one hand, and data users on the other hand, through technical, legal or other means, including for the exercise of data subjects' rights in relation to personal data" [40].

The term "commercial relations" is significant for this project. Although DataConsensus currently doesn't involve monetary transactions, future development might introduce such a mechanism. Thus, this legislation is relevant. Even now, it's essential to monitor potential regulations for data intermediaries to ensure the application aligns with best practices. Furthermore, DGA mandates neutrality for data intermediaries and prevents services like cloud service providers and data advertising brokers from becoming data intermediaries. This aims to increase competition and ensure the established players of the data economy today will not dominate this new market in the future [99].

In addition to data intermediaries, the DGA proposes a framework for data altruism organisations, enabling individuals and companies to share their data for "general interest" voluntarily and without reward. According to DGA, data altruism organisations must be non-profit and protect the rights and interests of the citizens or companies who share the data. Similarly to the data intermediaries regime, there will be certain rules to follow for data altruism organisations and they must register with the EU [22].

The legislation also briefly mentions data cooperatives, a core concept to this research and which is discussed further in Section 2.3.2. The Act considers data cooperatives as neutral intermediaries and acknowledges that they seek to balance the relationship between data subjects and data users. However, the DGA doesn't clarify the legal form and type of organisation that data cooperatives should take. Most importantly, the DGA does not provide guidance on the collective exercise of data subjects' rights under GDPR [13]. This will be further explored in the next section.

2.2.6 Critiques of the Data Act and DGA

Despite being separate legislation, the Data Act and DGA enable each other and they share many criticisms. Both legislation aim to encourage data sharing within the EU but commentators argue that it doesn't do enough to enable alternative data governance models. In particular, both Acts maintain GDPR's individualistic status quo, rather than expanding it to include collective stewardship over data [94], [77]. This means that there is still legal uncertainty surrounding the alternative data governance models, particular

data cooperatives [15]. However, as these regulatory frameworks develop, more clarity and improvements should occur.

Another significant argument against the Acts is that they prioritise the sharing of data for private profit, leading to further data commodification [94] [13] [112]. This critique overlooks issues around incentives that come with sharing data solely for altruistic and societal good purposes. There are situations where consumers and businesses alike will share data without any monetary exchange. This includes the current use case for the DataConsensus application, but allowing for compensation should incentivize more individuals in sharing their data and further normalise the concept.

In fact, the lack of profit incentive is a concern for data altruism organisations, commenters argue that the requirements for these organisations will be too onerous and in addition to their non-profit status, the legislation is likely to create a sustainable ecosystem for data altruism organisations [109]. Furthermore, the data altruism concept in the Act is weakly defined, particularly with regards to "general interest" and how it interacts with the GDPR's definition of "public interest".

As a whole, the legislation isn't as transformed as its initial ambition envisioned. Key rules and structures of the data intermediaries and data altruism organisations have yet to be established, though the Data Act will promote interoperability and easy access to data. It is the step in the right direction and highlights the EU's willingness to experiment on alternative data governance models.

2.3 Data Governance and Data Commons

In the section, we will first explore different frameworks to understand data governance, outline certain conceptual components and categorization to help understand the models of data governance discussed in relation to this project. Following this, we will discuss data commons and collective governance. In particular, this section will discuss the two data governance models most relevant to this project - data trusts and data cooperatives. To conclude, there will be a review of existing health data cooperative and other similar data sharing initiatives.

2.3.1 Understanding Data Governance

Data governance lacks a standard definition but can be described as "the exercise of authority and control over the management of data" [70]. It involves a web of actors and aims to maximise the value of the data while minimising the cost and risk of that data. While most definitions of data governance in the literature come from a corporate per-

spective, [11] proposes a more versatile but comprehensive definition - "Data governance specifies a cross-functional framework for managing data as a strategic enterprise asset. In doing so, data governance specifies decision rights and accountabilities for an organisation's decision-making about its data. Furthermore, data governance formalises data policies, standards, and procedures and monitors compliance."

This definition is broad but covers the core elements of data governance. Within this project, the framework, decision rights and data policies were the main focus but as this definition highlights, in a real world application, data governance covers much more. [11] further proposes a conceptual framework on data governance involving antecedents, mechanisms and consequences. Internal antecedents may be strategic or cultural, while external ones relate to laws and regulations. The mechanisms cover domain scope, organisational scope, and data scope, leading to various consequences.

[66] used Abraham's conceptual framework to establish 5 analytics dimensions of Data Governance: Stakeholders, Governance goals, Value from the data, Governance Mechanisms, and Reciprocity. The stakeholders reflect the many different actors involved in data governance, such as data subjects, data holders and data controllers. The governance goals are the different values and objectives that the stakeholders hold, such as enabling health research or preserving privacy. In this project, the value of the data depends on its usefulness on the research being performed.

Governance mechanisms is a commonly discussed concept within data governance literature and are the key tools for actors to implement their governance goals [42] [11]. Governance mechanisms can be better exploited by certain actors [11]. Recognizing this, [66] included the measurement of reciprocity or power relations between stakeholders as one of the key dimensions in their proposal. Designing thoughtful governance mechanisms that manage power relations is crucial to building a fair and sustainable governance model.

The above dimensions can be used to categorise different data governance models. Another categorization method is proposed by [110]. They approach the challenge from a data sharing perspective, and focus on two variables: engagement - "the degree to which the data supply and demand actors co-design the use of corporate data assets" and accessibility - "the conditionality of accessing private data by external parties". Specifically for engagement, a data holder might have no direct involvement over the data they are sharing, some involvement or they may be completely directing the use of the data. As for accessibility, the data is either open access or restricted access, meaning that pre-select partners to receive access. Based on this description, the accessibility variable is a simplified version of the Data Spectrum proposed by the Open Data Institute [76]. While this methodology in its current mode is relevantly simplistic, it does highlight the wide range of data sharing techniques.

2.3.2 Data Commons In Theory

Data commons, also known as data coalitions or collectives, is an emerging concept within the literature, proposed as an alternative to the dominant data governance framework of individual rights and responsibilities. Unfortunately, since data common is still an emerging term, there is no established definition yet within the literature. In fact, while certain literature treat data commons as a specific data governance model implementation, this work will treat data commons as an umbrella term to reflect any data space "that are collectively stewarded and governed by a community" [100]. Furthermore, a data space is an "infrastructure that enables data transactions between different data ecosystem parties based on the governance framework of that data space" [43]. A data commons is a data space, but not all data spaces are data commons.

Data commons encompasses a wide range of alternative data governance implementations, including open access commons, data cooperatives and data trusts. The latter two implementations will be explored in Section 2.3.2 and 2.3.2. Each implementation has its benefits and limitations, depending on the data involved, stakeholders, as well as the external antecedents, such regulation. However, they do share key characteristics and [100] proposes that the three pillars that contribute to these Data Commons implementations. The pillars are:

Stewarding Access This is similar to the accessibility variable proposed by [110], in which different data commons have different levels of access. The level of access is usually a balance between optimising the value of the data while preserving rights and minimising risks. The key governance mechanisms involved in stewarding access includes technical architecture, such as APIs and interoperability standards, permission interfaces and privacy-enhancing technology, such pseudonymization.

Collective Governance There are three components of collective governance outlined by [100]. The first is a defined community - while a data commons community doesn't have to be a traditional community, it is necessary to have some shared characteristics and shared values. Democratic control is another component of collective governance. There are varying degrees of democratic participation, from advisory boards, citizen panels and voting. This allows communities to have greater autonomy and it can also ensure accountability of other stakeholders. With that in mind, the final component is a trusted institution. Given the limitations of grassroot organisations, a trusted institution sustains a data commons, enabling the communities decision-making through its work, infrastructure and resources. In this project, the trusted institution would be the DataConsensus application.

Public Value The final pillar of a data common refers to the need for data commons to create products and services that benefit society as a whole. This is related to [66] dimension of value of data, however, [100] specifically argues that the data commons emphasise benefits for everyone and not just those managing the data. This pillar is arguably the weakness proposed because not all stakeholders will be motivated by the common good and in fact, for widespread adoption, incentives such as monetary compensation might be necessary. As such, this pillar is pretty inflexible and ignores other applications of data commons. Potentially a more flexible pillar would be the mission of the commons should be clearly defined, whether that be public good or private interests. Regardless, there are definitely many use cases where a data commons should be focused on the public good, which is the case in the DataConsensus application.

Benefits of Data Commons

The clash between data commons and today's dominant data governance model can be simplified to the collective vs individual approaches. In the current GDPR regime, the individual has rights but also responsibilities to manage their data. This puts a burden on the individual to engage with an onslaught of data requests and consent agreements. The literature has noted this and the resulting apathy and "consent fatigue" [101], where data subjects blindly agree to policies and this can very easily lead to poor data decisions. In fact, one impact of GDPR is so called "consent fatigue", where the Data commons can alleviate this burden, as the due diligence required is spread out across all the members of the data commons, not just the individual.

Another benefit of data commons is that they create a network effect, making it easier for third parties and recipients to process the data [82]. In a practical sense, this means that researchers will have access to large pools of data and this can accelerate and increase research. It also means that researchers will have a smaller burden to getting that access - instead of reaching out individually to each data holder, you only have to request from the data commons. Therefore, data commons would be a significant boon to the research industry and is in line with the objectives of the EU's data strategy.

Data Trusts

Data trusts are defined by the Open Data Institute as "a repeatable framework of terms and mechanisms that is mandated for use (or subject to scrutiny, or certification) in particular contexts to provide oversight of data access." [49] However, a more simplistic description of a data trust is that they give people a way to express their goals for the use of their data and require a trustee to work towards these goals. The structure of a data

trust matches a traditional common law trust, in which one party (a trustee) manages the rights associated with an asset for the benefit of another (a beneficiary) [82]. Given this and the other established norms for traditional trusts, adapting this framework to data is straightforward in theory but it is more complicated in practice.

The key difference between a data trust and a data cooperative is the trustee and the concept of fiduciary duty. The trustee is the trusted institution that the data commons relied on for its collective governance and the trustee is distinct from the data subject-s/holders of a data trust. Because of the legal structure of trusts, trustees are bound by a fiduciary duty or obligation of undivided loyalty to the beneficiary, therefore they are not allowed to act against the beneficiary's interests. This creates a mechanism for accountability towards the stewardship of the data, if the trustees abuse their position or are simply neglectful in protecting the data subject's privacy, the beneficiary/data subject has a legal course of action [82].

A caveat to the literature surrounding data trusts is that not all examples presented are legally considered trusts [49] [60] [108]. [82] suggests that either a contractual framework or an agent framework could be an alternative to a legal trust. However, legal trusts provide the greater accountability mechanism. Since this is a developing area, there are only a few examples of data trusts in practice, many of them piloted by the Open Data Institute in the UK and involve public bodies and citizens, though Alphabet also proposed a trust for its Sidewalk Labs in Toronto before it was abandoned [60]. Nearly all the literature argue for greater guidance and legislation to formalise these frameworks and protect users involved in these data commons.

The exact participation and involvement of data holders in a trust will depend on the practical implementation. This participation can range from systemic input from data holders to a hands off approach where trustees are simply directed by an pre-existing list of values and goals and are responsible for all other decisions [82]. Taking the latter approach, data trusts could share many characteristics with the democracy focused data cooperatives which will be covered in the next section.

Data Cooperatives

Data Cooperatives are a decentralised bottom-up data governance model in which data subjects 'voluntarily pool their data together, to create a common pool for mutual benefits' [50]. Members of a data cooperative are part-owners of the cooperative and all participate in the decision-making processes. Much like a data trust, data cooperatives originate from traditional governance frameworks, in this case, the cooperativism movement. Interestingly, the first Irish law regulating cooperatives is being drafted as of August 2023,

though it understandably doesn't make any mention of data cooperatives in the drafts, given how new of a concept it is. Despite the lack of legislation, it is opposite to establish a cooperative in Ireland as a company with a cooperative ethos constitution [45].

Data cooperatives share many of the same benefits of data trust and in fact, empower the data subject more than a data trust. Aside from the focus on democratic participation, the structural difference between a data cooperative and a data trust is that while a data trust uses the fiduciary duty to restrict the agenda of the trustee, a data cooperative has a more positive mechanism. Because the main stewardship force comes from the data holders/data subjects themselves in a cooperative, there is no need for fiduciary duty and the members can be more proactive in their agenda [82].

However, this presents a number of challenges. One challenge is representing the interests of all members. This highlights the necessity of a defined community as part of the [100] framework, having a defined community whose interests are similar and values are shared, eases the burden for a cooperative to find consensus and operate smoothly. This is a challenge that this work is looking to alleviate.

Interestingly it's not necessarily true that the power balance with a data cooperative is distributed equally among members. The constitution and rules of a cooperative could skew voting rights towards members who share more data or members who are part of the cooperative for the longest. However, within this project, all members have equal say. Even without a structural biassed constitution, there are ways that the democratic control leads to power imbalances between members. One example of this is where the democratic participation required for a data cooperative is too burdensome for all but the most motivated individuals [82]. This can be alleviated by creating a straightforward process with a friendly and accessible interface for decision making, which this project will endeavour to do.

Additionally, as data cooperatives grow and scale, the need for a level of trusted administrators grows. This is a significant issue that needs to be addressed and planned for when data cooperatives are being planned. There are similarities between administrators and trustees but crucially, an administrator should be a data subject and member of the cooperative as opposed to a trustee [67]. While this is a large area to address, this work won't dive into the specifics of administrator frameworks and governance.

Another issue with data cooperatives is funding, as they often rely on philanthropy. One solution is a membership fee [67], however this can discourage new members from joining. Another solution is a fee for data sharing access [53]. This project's application doesn't facilitate any financial exchange but it would be a next step if the application was going to sustain a long term data cooperative.

While data cooperatives have a number of challenges, the democratic aspect of it makes

it ethically interesting, since data subjects still have control over data decisions, even if it is collective control. Additionally, based on this project's analysis, data trusts and data cooperatives have many shared characteristics - a data trust can have a very strong democratic process for beneficiaries, while a data cooperative might have an administrator with some fiduciary duties to the other members of the cooperative. Therefore, while this work is focused on data cooperatives, many of the points are applicable in both.

2.3.3 Data Sharing In Practice

There are four notable organisations that are noteworthy for their advocacy of privacy-preserving data sharing. These include MyData Global, International Data Spaces Association, DataCollaboratives.org and OpenFuture. These organisations have produced any number of relevant concepts and guidance for data sharing that is relevant in this project and as such we will discuss them below.

MyData Global This international non-profit aims "to empower individuals by improving their right to self-determination regarding their personal data". Along with other literature, they have produced the MyData Declaration, which outlines six principles for a human-centric approach to personal data [85] These principles inform key aspects of the DataConsensus, such as transparency and interoperability. Additionally, MyData published a number of real-world case studies related to responsible data sharing, including one that inspired the defined community DataConsensus is built around [51].

International Data Spaces This association of over 140 member organisations aims to promote a technical standard for the data economy, motivated by the EU's Data Strategy. It proposes the IDSA's framework and the International Data Spaces Reference Architecture Model (IDS-RAM), as well as other literature and white papers related to data spaces [55]. The IDS-RAM utilises BPMN diagrams to outline its processes and this technique is utilised by DataConsensus to illustrate this proposed decision-making process [56]. However, it should be noted that IDSA is mainly driven by industrial data and has only recently been exploring personal data sharing [107].

Open Future Foundation This think tank advocates the open flow of data in the form of public data spaces and data commons. It performs research into interoperability, democratic data governance and how AI is changing data. Furthermore, it provides analysis and track key legislation and actions from the EU in the area of data spaces [78]. Literature from this organisation was invaluable in understanding the overall ecosystem of data commons.

Much of the existing literature has argued that data cooperatives can facilitate health research [82] and already there are existing data cooperatives that let members share their health data for scientific research. Section 2.3.3 will explore these examples and Section 2.3.3 will then explore other data sharing initiatives of note.

Existing Data Cooperatives

Midata Coop Founded in 2015 and based in Switzerland, Midata Coop is described as "a data trust organised as a co-operative" by a member of its management committee [73]. Interestingly, you can share your data through Midata projects without being a member of the cooperative. 20000 people have shared their data with Midata but the number of members is unclear. It should be noted that to become a member a one-time fee is required. This demonstrates the data trust characteristics of the organisation, since the members are similar to trustees for the non-members.

In terms of governance, there are in fact four bodies:

1. The General Assembly - This is the supreme governing body, though its legal powers are relevant high-level including electing the directors, auditors or changing the rules that govern the entire cooperative. Every member can participate and vote in this assembly and they must convene at least once a year.

2.

3. The Board of Directors - A group of at least 5 members, they are elected every two years and are responsible for overall supervision and leadership of the cooperative.

4. The Auditors - An independent entity whose duties are determined by law

5. The Council of Data Ethics - their responsibilities include reviewing the ethical quality of services and research projects and to make recommendations to the directors

ased on this structure, the cooperative pursues representative democracy, as opposed to direct democracy. Additionally, the different bodies serve as a check on the powers of the other bodies. The principles, rules and responsibilities of the cooperative are clearly outlined in its principles of association [67]. It should be noted that the exact process or criteria for how research projects are selected is unclear.

As for their technical architecture, Midata has developed a secure data platform which third parties can develop mobile apps on top of. These mobile apps then allow citizens to collect and share their health data in research projects. Interestingly, they argue that this method allows ordinary people to contribute to research as "citizen scientists". The

Midata platform is open source [69] and it uses structured personal data in Fast Healthcare Interoperability Resources (FHIR) format using SNOMED and LOINC ontologies [68]. This allows for easy exchange of healthcare-related information.

Salus Coop Salus Coop is a non-profit cooperative founded in Barcelona and operating since 2016. The research provided data includes biomedical, health and/or social studies. The data provided is also pseudonymised by Salus, before it's shared with a researcher. The research it provides data to researchers is required to be accessible to the public. Additionally, members can cancel the access to their data at any time [90].

The framework that Salus proposes is one where a data cooperative operates as a middleman between data keepers, the entities that own the databases storing citizen health data, and data users, the parties who are interested in accessing the cooperative data for research. The data flow of the cooperative can be seen in Figure 2.2. Based on this figure, you can see that the cooperative doesn't actually own the data directly but must interact with the data keepers (as opposed to the data subjects) to share it. Unfortunately, there is limited information on collective decisions and the process involved [53].

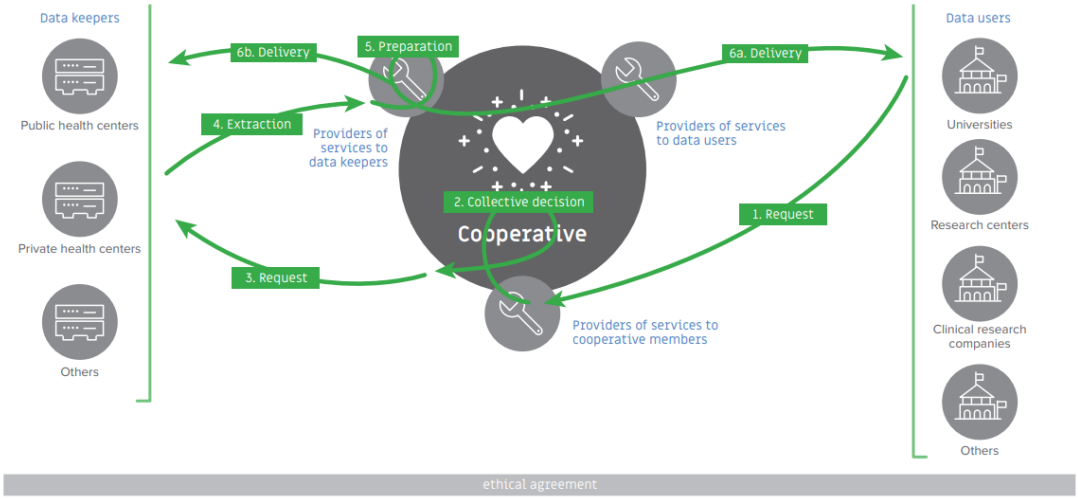


Figure 2.2: The data flow of the Salus Coop, sourced from [53]

In terms of the technical architecture, the Salus Coop utilises the blockchain to store its registry of agreements but the format of these agreements is unspecified. Likely Salus Coop will adjust these agreements with any standards proposed by the EU for smart contracts in the coming years. As for funding, the cooperative has revenue streams from payments from the data users and membership fees. The payments from the data user varies depending on the case [53].

Other Data Sharing Initiatives

There are numerous data sharing initiatives between groups of data subjects and researchers already existing in practice that don't fit into the data cooperative mould. OpenHumans is one example of these initiatives, allowing individuals to import, pseudonymised and share their personal data with the projects that they choose [60][46]. However, users can also share their data once with a data commons which will facilitate data sharing with multiple studies and projects. One example of these data commons is the NightScout Data Commons [104]. Projects connected to a data commons must be approved by the members of that data commons though the exact decision process is unclear. Finally, the funding of OpenHumans appears to be mostly philanthropic, and there doesn't appear to be any fee for projects to access user data.

One form of data sharing initiatives that are well established are patient registries - "secure, centralised databases, containing health data on people with a specific diagnosis or condition" and these registries can be used by approved researchers [18]. Participation is voluntary for patients. Within Ireland, patient registries are generally operated as charity and managed by a board of directors. One example this section will explore is the Cystic Fibrosis Registry Ireland (CFRI). Its board has one patient representative and the rest are medical professionals. Given its structure, medical professionals are the driving force behind patient registries but there are obvious benefits for patients too.

To gain access to the CFRI data, the research must be approved by the CFRI Scientific Research Committee and there is a short fee to make an application. Notably, the CFRI generally only shares summary statistics and rarely releases raw anonymised patient data. However, CFRI does contribute anonymised data to the European CF Society Patient Registry (ECFSPR), which equally allows researchers to request data. Given this international collaboration, there is already a body of work to modernise patient registries into RDF data structures to allow for more harmonisation and interoperability [64].

These alternative data sharing initiatives demonstrate that there is a clear interest in sharing health related data. However, OpenHumans suffers from a lack of transparency and clear governance, making it difficult to attract new members, while patient registries are not driven by the data subjects themselves but by the data users. The gaps can be fulfilled by data cooperatives.

2.4 Solid and Linked Data

This section will examine the background and related work of Solid and privacy-related policy languages and vocabularies. Section 2.4.1 examines Solid, the technology that

makes the cornerstone of this project. Following this, Section 2.4.2 delves into linked data, which is the key tool for ensuring interoperability in this application. Specifically, it explores the privacy related policy languages and vocabularies.

2.4.1 Solid

[80] describes solid as "an ongoing effort to decentralise the control of data by moving its storage away from centralised systems and into Pods that are controlled by individuals" . Solid specification was first released in 2016 and was developed initially at the Massachusetts Institute of Technology and led by Sir Tim Berners-Lee, the inventor of the World Wide Web. The specification defines the interaction model for how Pods manage identity, access and communication, based on existing web standards. The core principle of Solid is data sovereignty where individuals control their data and how it is used and as such, Solid has gained interest from privacy-conscious developers as an alternative approach to the centralised lock-ins of Big Tech platforms [95].

Pods, which stands for Personal Online Datastores, can be self-hosted or managed by a pod provider. To interact with your pod, solid provides a HTTP RESTful API. With the correct access, applications can easily create, read, update, and delete data (CRUD operations) in a Pod. Pods use existing cloud technologies and servers similar to traditional data management approach but what differentiates the Solid specification is based on Linked Data. Linked data is a broad term but refers to best practices for publishing structured data on the Web and it is linked to the Semantic Web, a vision for a machine-readable web [8]. The greatest benefit of Linked Data is interoperability, which in the case of a decentralised data network such as the one Solid envisions, is necessary.

Hand-in-hand with Linked Data is the Resource Description Framework (RDF), a standard structure for expressing information about resources in a way that preserves meaning regardless of the application. A resource can be anything, from documents, living beings or conceptual ideas. Another component of RDF is the triples, which is an expression in the form of subject predicate object. Given the standardised approach to expressing data, RDF is the basis for the policies used in the DataConsensus application, as well as its other data structures.

Authentication

Another component to understand in relation to Solid is the WebID. A WebID is an Internationalised Resource Identifier (IRI) and is used to identify an Agent (a person, organisation, or software). A user links their sharing preferences to the WebID of a third party in order to share data with them. To verify the identity of agents, the Solid

OpenID Connect (Solid OIDC) specification is used and authentication is performed by an OpenID provider [20]. The benefit of this is that solid applications don't need to store sensitive user credentials or manage the authentication process themselves. An example of a WebID is https://id.inrupt.com/odonneb4.

Authorization

There are two specifications that govern the authorization protocol of Solid - the Web Access Control (WAC) and the Access Control Process (ACP). A solid pod may conform to one or both of these specifications so a solid application must comply with both. Fortunately, inrupt's solid-client javascript library has a universal access control APIs that can be used with either specification. The universal API allows for three access modes: read - the ability to view the contents of a Resource, append - the ability to add new data to a Resource, write - the ability to add new data to a Resource, and to change or remove existing data [1].

Both specifications also produce records for access granted to an app for specific data in either a resource's Access Control Lists (ACLs) for WAC or resource's Access Control Resource (ACR) for ACP [6][5]. However, these records are "not sufficient to meet GDPR's requirements for either valid consent or a valid record of consent". Neither WAC nor ACP cover required information by GDPR such as purpose, legal basis and other requirements discussed in Section 2.2.1. This gap in the solid's functionality has prompted research into integrating the more expressive machine-readable policies into solid, which will be explored further in Section 4.2.

Applications

There are numerous applications for the Solid protocol and pods. [26] reviews potential examples and user journeys for data sharing within Solid. It evaluates these examples with a set of decision tables to determine which patterns are legally valid. The examples used are not exhaustive but the paper does demonstrate the basic functionality of solid and the legal considerations to keep in mind. One interesting question proposed by the report is "where should the consent be stored?". Since a copy of the consent must be given to the data subject in the cases of health research, as discussed in Section 2.2.2, there is a question over whether pod access to the consent statements (aka this application's agreements) is enough to ensure compliance or whether each user must have a separate resource for their consent statement in their individual pods.

2.4.2 Privacy Related Policy Languages and Vocabularies

Within the realm of linked data, ontologies express concepts and the connections between them in ways that are both machine- and human-readable. Ontologies are used to model a specific domain and provide a shared understanding of that domain. In the case for this project, relevant ontologies model domains related to privacy, legal data requirements and consent. Furthermore, vocabularies refer to a set of terms used to describe concepts within a specific domain. Unlike ontologies, vocabularies are less concerned with defining relationships and rules and more focused on providing a standardised set of terms. Finally, Policy languages are formal languages used to express rules. Machine-readable examples include XACML5 (eXtensible Access Control Markup Language) and ODRL (Online Digital Rights Langauge) [61].

In this section we will discuss the relevant policy languages, ontologies and vocabularies. Work from [32] has reviewed 13 privacy-related policy languages and 9 data protection vocabularies. A useful contribution from this work is a breakdown of the informational elements necessary for the management of each GDPR right and obligation. Using this table as an evaluation tool, they concluded that ODRL and LegalRuleML, when complemented with DPV and GDPRtEXT, are the most holistic languages and vocabularies to fulfil the most information requirements for GDPR.

Additionally, a body of work has already been done using ODRL to describe policies relating to the use of personal data, namely those linked to the GDPR [12] [27]. [61] reviewed the consent models of existing vocabularies, including DPV. It notes that DPV is based on the requirements of GDPR for recording and documenting consent. Of the 16 competency questions used to evaluate the vocabularies, the two that DPV failed to fulfil were "When was consent given/revoked?" and "What are the responsibilities of Data Controller?". Interestingly, by integrating DPV with ODRL, these limitations could be overcomed.

ODRL

The Online Digital Rights Language (ODRL) is a policy expression language and a W3C standard, providing a flexible and interoperable information model and a vocabulary of terms. ODRL provides a structured framework to express policies for different use cases, such as software licences to access and usage control policies. An ODRL policy is a group of rules. ODRL rules can be broken down into permissions, prohibitions and/or duties. The most relevant rule type in this project is a permission, which is the ability to exercise an action or an operation over an asset. Another key aspect to a policy is the Constraints which are the conditions or restrictions placed on actions [52]. A constraint has three

main elements:

- leftOperand - defines what aspect of the asset the constraint is applied to;
- operator - defines the relationship between the leftOperand and the rightOperand;
- rightOperand - defines the value or set of values to which the leftOperand is compared using the specified operator.

The ODRL vocabulary only supports a minimally policy expression so to adapt the policy to different domains, other ontologies and vocabularies can be incorporated into an ODRL policy. To do this, an ODRL profile must be defined, which defines specific sets of classes and concepts.

DPV

The Data Privacy Vocabulary (DPV) is a vocabulary and ontology for expressing information related to use and processing of personal data in RDF machine-readable data. The DPV has a number of companion vocabularies, extending its coverage of technology, GDPR and other relevant concepts based on privacy and data protection. This effort is part of its ambition to be globally useful vocabulary to be used across all domains where privacy may be involved. The core taxonomy of DPV covers key concepts and definitions in GDPR, including purpose, personal data, legal basis and key legal roles [79]. Because of these qualities, DPV is an ideal vocabulary to integrate into any Linked Data application with a focus on data control or privacy.

OAC

The ODRL profile for Access Control (OAC) defines access control policies for Solid using DPV. OAC serves as the policy layer of the Solid protocol below the access authorisation layer and would provide more granularity in the access control mechanism. One objective for the OAC is to allow users to express their privacy preferences and requirements over specific resources within a pod and those policies can be matched with access requests, a mechanism described as automated consent [30].

OAC includes the following as its core concept: Preference Policy, Requirement Policy, Personal Data, Processing, Purpose, Recipients, Legal Basis and Technical and Organisational Measure [31]. These concepts relied on vocabulary from DPV and for this project, the concepts related to ODRL constraints are particularly noteworthy and will be expanded upon.

Whether OAC signifies consent in a significant question within the literature. [41] explores the question specifically in regards to automated consent. It finds that more legal research is needed to confirm "to what extent [automated consent] can be expressed by individuals with the help of technologies, such as the Solid protocol". It does explore the requirements to express consent, including specificity of purposes, disclosure of the identity of the data controller and third party recipients, as well as special requirements of biomedical research. While this project doesn't focus on automated consent, the discussion of the disclosure of identity is noteworthy as the basic identity expressed by an ODRL policy is a WebID which is unlikely to pass GDPR compliance. Therefore, the final data sharing tool must take that into account.

2.5 Summary

This project required a multidisciplinary approach, where technical knowledge is combined with legislative and governmental knowledge. The legislation was examined in order to extract the relevant requirements, as well as understand the data sharing ecosystem will change in the future. A look into data governance highlights the importance of the different actors and governance mechanisms in a data governance model. Additionally the examination of the alternative data models and their practical implementations justified the benefits of the project, while illustrating the challenges this work and future work will face. The final section provides a technical background of solid and the linked data technologies used in this project. This provides a foundation to understand the technical implementation of this project.

The key conclusions to draw from the related work is that the Irish interpretation of GDPR for health research puts an onerous burden on researchers. Despite the Data Governance Act missing key opportunities to legislate for collective data governance models, the EU as a whole is striving to facilitate more data sharing. The literature backs the conclusion that data cooperatives can balance data subject rights and the benefits of data sharing. While existing data cooperatives involve representation democracy as opposed to direct democracy, their exact decision processes are obscure and a significant gap in the literature. In the exploration of the solid and linked data, this project concluded that solid is an appropriate technology to use since it matches the ethos of this project and facilitates interoperability through linked data. There is already a body of work utilising ODRL and DPV to standardise consent agreements and this project builds on this existing work to create a collective approach to data sharing.

Chapter 3

Design

This Chapter outlines the key design choices for the DataConsensus application, including the identified actors, requirements and functional architecture. Additionally, this chapter describes the defined community which the DataConsensus application is built around. The chapter also describes the ontology developed for this application, in particular the ODRL Profile for Collective Policies in Section 3.5.1. Furthermore Section 3.6 explains the decision process adopted in the DataConsensus application.

3.1 Defined Community

Chapter 2 introduced many potential and existing use cases for data cooperatives. While it is a goal of this project for its contributions to be interoperable and adaptable, a specific and realistic problem overview was desired for design and development purposes. Furthermore, as outlined in Section 2.3.2, a core component of collective governance is a defined community. Therefore, it is important when approaching any software development for data commons to identify a clearly defined community.

Given the Data Act's measure to provide individuals more access to the data they generate on their connected devices [23], wearables were explored as a potential source for this project's defined community. One wearable identified with a dedicated group of users were continuous glucose monitors (CGM). These are generally used by diabetes patients to track their real time blood sugar levels.

A community with a shared diagnosis was desired for this project, since surveys have shown that people living with a rare disease are very likely to share their health data to improve research into their own disease as well as others [38]. While diabetes is not classified as a rare disease, it can be assumed that patients with diabetes would still be more likely to share health data compared to healthy individuals. This overcomes a major

issue faced by data cooperatives, as many potential members don't feel comfortable in sharing sensitive data, especially without compensation. However, there is an incentive for patients with chronic diseases, like diabetes, since any research into it or other related areas could improve their own treatment.

The most deciding factor for choosing patients with diabetes as this project's defined community was Nightscout. Nightscout is an open-source software project that allows CGM users to monitor their blood sugar levels remotely on their smartphone and to export that data onto the cloud [106]. Nightscout is a community-led initiative, initially developed by parents of children with Type 1 Diabetes [63]. Nightscout was released in 2014, it can currently work with the most common CGMs on the market and it has ample tutorials and user-friendly features. Additionally, the Nightscout software is supported by the Nightscout Foundation, a Texan non-profit [105].

It appears to be particularly straightforward to export Nightscout data in a standardized way [98] [87]. This is particularly ideal for this project, since having one standardised format makes pseudonymised and combining the data far simpler. Within this project, the exact data used from NightScout is its entities.csv, as seen in Appendix C.1. Since this project lacked a CGM device in order to create a NightScout account and experiment with the platform, no automated or integrated solution is provided for the DataConsensus application. Instead, users will be expected to manually export the entities.csv and upload into a solid pod.

Finally, as mentioned in Chapter 2, the Nightscout community has already participated in data sharing initiatives for research. The Nightscout Data Commons on OpenHuman was launched in 2017 and has 200 members [104]. Unfortunately, the Data Commons' webpage is currently defunct so it is unclear if the initiative is still active or how it was governed and operated. Regardless, this activity within the community demonstrates that there is some desire among the users to share their data with researchers. Furthermore, there have already been projects associated with MyData.org to develop the technical frameworks to easily share Nightscout CGM data with healthcare professionals [92] [51]. As for Nightscout's adoption in Ireland, there are no public figures but a private Facebook group called Nightscout Ireland currently has 259 members [9] so there should be potential members in Ireland to recruit.

Given the adoption of Nightscout, its existing governance structure with the Nightscout Foundation and established community, along with the technical expertise of the developer community of Nightscout, they appear to be an ideal community to design the DataConsensus application for. Therefore, the DataConsenus application will be designed with this community in mind.

3.2 Actors

The actors of the DataConsensus application were extrapolated based on the literature review performed in Chapter 2, incorporating roles and characteristics from both legislation and Data Commons literature. There were three actors identified: Member, Third Party and Admin.

3.2.1 Member

A member in this context refers to a data subject whose data is part of a larger pool. Since the data is their own personal data, they fall into both legal categories of data subjects and data holders. Additionally, the members make up the defined community of DataConsensus. Following the naming process of other data cooperatives, this type of actor is labelled a member since they participate in the decision making process. Members within this project are users of CGMs and likely part of the NightScout community.

In regards to key interactions with the application, members have the ability to vote on requests and offers. They can also submit their own offers for consideration and contribute to the policy discussion by adding comments. Within this report, the group of members that pool their data will be referred to as either the Data Group or the Data Community.

3.2.2 Third Party

A third party is a person or entity, such as a researcher, looking to access and use the members' pooled data. With regards to the legislation, a third party falls into GDPR's "third party" definition, as well as the Data Act's "data recipient" definition when they have been granted access [39] [36]. The third party actors in this work are similar to the data users (Salus Coop) and partners (Midata Coop) actors in other data cooperatives [53] [68]. Third party was chosen as the name of the actor, since it is more distinct.

Within the application, a third party will be able to submit requests for data access. They may also agree to an offer from the members if their initial request is voted down. If access is granted, the third party will access resources in the system's pod that have been collected from the users and pseudo-anonymised.

3.2.3 Admin

Based on the governance structures of existing data cooperatives, as described in Chapter 2, it was decided to add an admin actor within the application. The main objective of the admin is to oversee the system and prevent any abuse within the system from bad actors.

These bad actors include members and third parties. Especially as the system scales, an admin layer of governance appears to be necessary and therefore having the functionality and the structure already there will be beneficial.

In terms of permissions and functionality, the admin has the final approval before an agreement policy can be executed and access is granted to resources. However, they will not be able to enforce a policy or access without the approval from members. Admins can revoke agreements. Therefore, the system is designed to favour less data-sharing over more. Admin also set the rules surrounding deliberations, including time frames and thresholds. Furthermore, they have the authority to moderate user comments to maintain a respectful and productive environment.

It should be noted that the DataConsensus application in its current state is lacking functionality for electing or appointing admins. Given the limitations of this project, it was decided that it was out of scope. Nevertheless, future work into providing this functionality and reviewing the appropriate governance mechanism for admins is welcomed and encouraged.

3.3 Functional Requirements

This section presents the requirements for the DataConsensus application. These requirements take the form of three user stories, as described in Section 3.3.1, as well as the use cases which are outlined in Section 3.3.3.

3.3.1 User Stories

The software requirements were developed using three user stories, describing three scenarios for how the actors would interact with the software and each other. These user stories were developed based on limited real world information from data sharing initiatives. The first two stories are straightforward implementations of the system, while the third story illustrates the complex collaboration functionality between members and is the significant differentiator between other data sharing initiatives or ODRL profiles. The three user stories will serve as software system requirements in our evaluation.

User Story: Academic Researcher

The third party is a Principal Investigator (PI) at a respected academic research institution. They are leading a research project examining how glucose patterns of diabetes patients vary through the day. The DataConsensus data pool provides a straightforward way for the project to gain the glucose patterns data they need to complete the research.

The PI believes this research will provide more evidence-based recommendations for how people with diabetes should manage their glucose levels. The PI plans on publishing a paper on this research, which will also benefit the wider scientific community researching glucose and will likely lead to further research ideas.

In terms of the data request's purpose, it will be solely academic research and the PI plans to use the data only for this specific project. The research project will also not sell the data nor sell any insights from the data. The PI is working with a PhD student and a data scientist, both a part of the academic institution. As such, the PI, PhD and data scientist will all have access to the data in order to complete the research.

The academic research institution the PI is a part of mandates that all staff members take formal cybersecurity and data protection training. The institution mandates that all researchers follow an ethical code of conduct. The institution also promotes a Privacy by Design approach to any research project. The institution also employs a data protection officer, who has been consulted on this project. The data will be stored and transferred securely using encryption and access to the data will only be granted through password authentication. There are also compliance monitoring procedures and a plan in place for incident reporting if there is any unauthorised data processing. The data will only be processed in Ireland and won't be shared with any International collaborators [102].

Given the thorough constraints listed above, the trusted reputation of the organisation and the altruistic purpose of the research, the sentiment among the data group is very positive and the members overwhelmingly vote to approve the data request.

Once granted access by the application's users, the academic research institution accesses the pseudo-anonymised data from the DataConsensus pod. The data will then be analysed using statistical methods, machine learning algorithms, and other research techniques to identify patterns, associations, and trends. At the data agreed upon, any duplicate data files of the pod data will be destroyed.

User Story: Insurance Company

The third party is an Irish health insurance company. They are interested in accessing the CGM data to gain insights into glucose patterns and develop risk assessment models for their policyholders. They aim to analyse the data to identify patterns that could help them understand the relationship between glucose levels and health outcomes, particularly in individuals with diabetes. The insurance company states a number of benefits if the research proceeds. These benefits include more efficient and accurate underwriting for diabetes patients, more personalised insurance products, as well as incentives and rewards to policyholders for managing their glucose levels effectively.

The purpose for the data request is commercial research but they commit to not selling it with any unauthorised parties, though they may sell the insights. The company follows the guidance from Insurance Ireland and has stringent cybersecurity and data protection measures [54]. These include GDPR training and regular cybersecurity training for all its staff. The staff have also all signed NDAs regarding any personal data. All computer systems within their organisation require two factor authentication, as well as encryption. Additionally, the company also has a designated data protection officer responsible for overseeing compliance with privacy regulations.

Furthermore, the company employs encryption during storage and transmission, with industry-leading cryptography. There are also network security protocols and on-site physical security control. Regular compliance monitoring procedures and a certification system are in place to ensure ongoing adherence to data protection regulations. In the event of any unauthorised data processing or data loss incidents, the company has a well-defined plan for managing and reporting the incident [111].

The patient group carefully evaluates the request and deliberate on it amongst themselves. Comments posted on the DataConsensus platform below the data request include many negative ones such as "While the security measures seem good, I just don't know if I trust an insurance company with my health data. They definitely have ulterior motives"[24] [97]. While the constraints outlined above are stringent and well-thought out, a number of members don't trust the intentions of the insurance company and successfully convince a majority of members that with this data, the insurance company could increase insurance premiums for diabetes patients. Following this, the group voted on the request and they rejected the request.

3.3.2 User Story: Non-Profit Organisation

The third party is a newly established nonprofit organisation looking to improve healthcare resources for people with diabetes. Their mission is predominantly advocacy focused but they want to be data-led and evidence-driven. Therefore, they desire the blood sugar data to better understand the needs of diabetes patients. Still, the exact aim of their research is not stated and it appears to be mostly exploratory.

As they are a new organisation with limited resources, they have opted for more relaxed data protection measures. They do not have a designated data protection officer and do not have training or a code of conduct for their staff to follow. They do have standard operating system security and password authentication but cannot guarantee encryption at rest. They do commit to not selling the data or sharing it with unauthorised parties. They have also requested a very long duration of 20 years for the research.

This request prompts a significant discussion among the members of the group. Although concerned about data protection, many members believe in the nonprofit organisation's good intentions and their potential to make a positive impact. Therefore, one member proposes a counter offer. This offer outlines the existing constraints from the request and adds the following:

- The staff of the nonprofit must have data protection training.
- The nonprofit must follow an ethical code of conduct.
- The nonprofit must employ a data protection officer.
- The data must be stored and transferred securely using encryption.
- There must be a plan in place for incident reporting if there is any unauthorised data processing.
- The duration will be reduced from 20 years to 1 year.

The initial request is rejected during the request deliberation and following this, the counter offer is discussed and voted on by the group. The group approves of the counter offer and the system forwards it onto the requester. After reviewing the offer and adjusting their operations, the nonprofit agrees to the counter offer.

Once granted access by the application's users, the nonprofit accesses the pseudo-anonymised data from the DataConsensus pod. The data will then be analysed using statistical methods, machine learning algorithms, and other research techniques to identify patterns, associations, and trends. At the data agreed upon, any duplicate data files of the pod data will be destroyed.

3.3.3 Use Cases

Figure 3.1 illustrates the use cases for each actor within the system. While some of these use cases were identified through the user stories outlined in Section 3.3.1, others were added as part of the legal requirements (such as "withdraw consent") while others exist to improve the quality of life, such as the ability to view your profile and others.

3.4 Functional Architecture

This Section describes the functional architecture of the DataConsensus application. Figure 3.2 presents a diagram of this architecture and the individual components are following:

User Interface: This is the component of the DataConsensus application that users interact with. It is accessible from a web browser and dynamically changes depending on what type of user is viewing it.

User Manager: The user manager links the user interface to the Solid Pod Providers. As mentioned in Section 2.4, Solid provides an external authentication process using WebIDs, which is operated by the external Solid Identity Providers. However, the User Manager piggybacks on this process to identify existing users and their use type. This component is critical for managing users' access rights, authenticating user identities, and maintaining the security of user data.

Application Service: The Application Service acts as a processing hub for all the requests coming in through the interface. It receives HTTP requests from the interface, maps these requests to appropriate REST API endpoints, processes the requests and returns the response back to the interface.

Data Transformer: This component is responsible for the pseudonymization of the user's data to protect user privacy, and standardising the data into a single format suitable for storing in the Data Consensus Pod.

CRUD Service: CRUD is an acronym for Create, Read, Update, and Delete. These are the four basic functions for performing these operations on the data stored in the Solid pods, serving as an interface between the pods and the rest of the system.

Access Control: The Access Control component manages permissions for accessing resources in the application's pod. It has the ability to grant or revoke access and it ensures that the data in the application's pod is only accessible to the appropriate users.

Vote Calculator: This component is responsible for calculating the results of the request deliberations and the offer deliberations.

Mailer: This component provides the functionality to send mail notifications to users.

3.5 Ontology

This section will explain the ontology and schema developed for the DataConsensus application. The core component of the ontology is the ODRL Profile for Collective Policies

(OCP). Additionally, the rest of the data within the system, such as comments and votes, are structured as linked data and this will be described too in this section. There are 5 schemas to this work - ocp.ttl, project.ttl, user.ttl, comment.ttl and vote.ttl. Figure 3.3 presents the relationships between the classes defined in the schema. OCP will be discussed in depth below in Section 3.5.1. Table 3.1 shows the prefixes and namespaces for the 5 schemas developed for the application, while a table all namespaces used in this project can be seen in Appendix B.

The most important class outside of the OCP is `project:Project`, a subclass of `dct:collection`. A project groups the related policies together and stores relevant information that is shared across all of the policies. This includes the timeframe of the deliberations or the title of the research project requesting the data. The other crucial property of the `project:Project` class is the `project:hasProjectStatus`, which communicates what stage the project is at.

Prefix	**Namespace**
ocp	`https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schemas/ocp.ttl#`
user	`https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schemas/user#`
project	`https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schemas/project#`
vote	`https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schema/vote#`
comment	`https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schema/comment#`

Table 3.1: Prefixes and namespaces of the schemas used in DataConsensus

3.5.1 ODRL Profile for Collective Policies

The OCP was developed following the best practices proposed by the ODRL community. This project is currently investigating publishing a human readable specification as a HTML web document, along with requirements for registering a profile with the ODRL community.

Based on the user stories in Section 3.3.1 there are three policy subclasses in the proposed profile: `ocp:Request`, `ocp:Offer` and `ocp:Agreement`. A request is submitted by a third party, an offer is submitted by a member in response to a request. Finally, an agreement is created when all three actors - the third party, the members and an admin

approve of a policy. Once an agreement is created, access to the data pool is granted to the third party.

An example of each policy subclass can be found in Appendix B. OCP are building off of the policies proposed in OAC [31] and in the work of [91]. The three policy subclasses share the same properties, with the agreement subclass having an additional `dct:references` property that links the agreement to the policy that was approved by all three actors. Agreements within the system are automatically created and share the same constraints as the policy they reference.

The approval status for each actor is tracked by three properties: `ocp:memberApproved`, `ocp:adminApproved` and `ocp:thirdPartyApproved`. The range of these properties is the Status class which has the following subclasses: `ocp:Pending`, `ocp:Approved`, `ocp:Rejected`, `ocp:Blocked` and `ocp:Revoked`. It should be noted that the `ocp:Blocked` status subclass reflects when the approval of the actor is not needed because of the actions of other actors. The `ocp:Revoked` status occurs when the actor has previously approved of the policy but has revoked their permission, which allows the easy tracking of consent changes. The other status subclasses are self-explanatory. Additionally, the `ocp:memberApproved` can only be changed to `ocp:Approved` and `ocp:Rejected` when the members have voted during either the request or offer deliberations. The current iteration of DataConsensus does not have the functionality for members to vote to revoke a policy, instead individual members can leave the group if they no longer wish to consent to the project or the admin can revoke on behalf of all the members. Improving this functionality is future work of the DataConsensus application.

Based on the discussion from Chapter 2, there is specific information that should appear in the policies. These include legally required information, as well as information that aids the decision making process. The full list of extracted requirements will be discussed in Chapter 5 to evaluate the OCP.

The structure of the policies can be broken down into three types of instances: the main policy instance, the permission instance and the constraint instances. These instances are separated instead of combined as one nested instance because of the limitations of Solid. The main policy instance follows the same format of the OAC policies, with the additions of `dct:isPartOf`, `ocp:hasJustification`, `ocp:hasConsequences` and the approval status discussed above. For agreement policies, there are three additional properties, `dpv:hasDataSubject`, `dpv:hasJointDataController` and `dpv:hasLegalBasis`. The legal basis for all agreements is consent, in accordance with the HRR. A joint data controller approach is taken so it specifies that the third party and the webID of the DataConsensus application. The data subjects are all members of the data cooperative.

The permission instance defines the actions the `odrl:assignee`, the third party, can

take on their `odrl:target`. In all permission instances, the `odrl:assigner` is DataConsensus's webID and the `odrl:target` points to the URL of the data pool. Additionally, for all policies, the actions permitted to the third party are `dpv:Store`, `dpv:Use`, `dpv:Transform`, `dpv:Remove`, `dpv:Copy`. `dpv:Use` and `dpv:Transform` accommodate data access and analysis for the third parties, while `dpv:Copy` and `dpv:Store` depict the process of the third party transferring and retaining data from the Data Consensus pod. The addition of `dpv:Remove` underscores the requirement for the third party to delete the data once the project duration specified in the policy concludes. Future iterations of the DataConsensus application could allow for more customisation of this property, allowing users to specify the actions or even point to different pools of data.

Finally, there are the constraints. The four `odrl:leftOperands` defined in OAC [31] are included in these constraints. OCP goes further and defines the DPV terms of `Organisation`, `UntilTimeDuration`, `Jurisdiction` and `ThirdCountries` as `odrl:leftOperands`. Figure 3.4 compares OCP to OAC and ODRL. To allow `odrl:rightOperands` have multiple terms, OCP defines `odrl:isAnyOf` and `odrl:isAllOf` as `operators` too.

The DataConsensus policies have eight or nine constraints defined. These include the following:

Purpose Constraint: This constraint specifies the purpose of the policy, such as Academic Research.

Selling Data Constraint: Utilising the `oac:purpose` again, this constraint specifies whether or not the assignee can sell the data. Both the affirmative and negative specifications are equally vital in clarifying the policy, ensuring that there is no ambiguity in understanding the allowed and prohibited actions.

Selling Insights Constraint: Similarly to the Selling Data Constraint, this constraint specifies whether the assignee can sell insights gathered from the data, including summary data.

Duration Constraint: This constraint specifies the expiration time of the agreement, or the date when the assignee can no longer access or process the data. The third party will also be asked into a justification for the duration.

Organisation Constraint: This constraint specifies the type of organisation of the assignee, such as a Governmental Organisation.

Recipient Constraint: This constraint lists the recipients that may receive the data or the results of the processing of personal data. The third party will also be asked into a justification for the duration.

Technical Organisational Measure Constraint: This constraint specifies various technical and organisational measures to ensure a secure processing of the data. The third party, in submitting or approving of the policy, is confirming that that they have the arrangements in place to abide by these measures.

Jurisdiction Constraint: This constraint defines the legal jurisdiction for the policy, it currently defaults to Ireland.

Third Country Constraint: This constraint specifies whether the assignee will be sending the data outside the specified jurisdiction. This constraint is optional.

Vocabulary Choices

Given the wealth of vocabulary terms available within the DPV, a subset of more relevant terms were selected for ease of use and understanding between third parties and members. These terms were selected based on existing policies from patient registries and personal data research projects, including the Terms & conditions of data use for the Cystic Fibrosis Registry Ireland [17] and other literature gathered to support the user stories.

For terms related to organisations, 7 were selected and reflect a wide range of organisations to ensure all types of third parties are facilitated through DataConsensus. Regarding purpose, 5 terms were selected, though there are 78 terms related to purpose available through DPV. Three terms reflect specific types of research - `dpv:AcademicResearch`, `CommercialResearch`, `dpv:NonCommercialResearch`. There are also the additional purposes of `dpv:SellInsightsFromData` and `dpv:SellDataToThirdParties`, utilised for their own constraints.

There are 29 Technical and Organisational Measure terms selected from DPV for the application. Table 3.2 groups and justifies the inclusion of specific terms.

3.6 Decision Process

A key contribution from this work is a BPMN diagram outlining the decision process involved in the DataConsensus application. Figure 3.5 presents this diagram and the BPMN XML can be found in this project's GitHub repository: `https://github.com/bod777/DataConsensus.git`.

Term	Justification
dpv:OperatingSystemSecurity dpv:NeworkSecurityProtocols dpv:CryptographicMethods dpv:EncryptionInUse dpv:EncryptionInTransfer dpv:EncryptionAtRest dpv:EndToEndEncrption	These terms were selected as they are the key concepts when it comes to encryption, which is a commonly mentioned measure for data sharing agreements. The range of encryption terms included allows for some amount of specification and customisation.
dpv:MultiFactorAuthentication dpv:PasswordAuthentication dpv:SingleSignOn dpv:UsageControl dpv:PhysicalAssessControlMethod	These terms relate to the common techniques to police access to data.
dpv:ProfessionalTraining dpv:CybersecurityTraining dpv:DataProtectionTraining	These terms represent the common training provided to researchers and people with exposure to sensitive data.
dpv:CertificationSeal	This term was included as certificates are a growing standard to prove GDPR compliance. In future iterations of this application, this term could even be a odrl:leftOperand for users to specify their exact certification.
dpv:NDA dpv:CodeOfConduct dpv:ConsultationWithDPO dpv:dpia	These terms were included as they are considered required information for researchers to include in their requests.
dpv:DataProcessingAgreement	This term reflects the inclusion of external agreement outside of the DataConsensus system.
dpv:PrivacyByDefault dpv:DesignStandard	These are design approaches promoted by established research institutions.
dpv:AssetManagementProcedures dpv:LoggingPolicies dpv:MonitoringPolicies dpv:ComplianceMonitoring dpv:IncidentManagementProcedures dpv:IncidentReportingCommunication dpv:ReviewProcedure	These terms reflect the common procedures involved in data protection measures.

Table 3.2: The technical and organisational measures vocabulary included in the DataConsensus application, along with the justification for their inclusion.

Because of the lack of information regarding the decision-making processes for existing data cooperatives, this publicly available process may be useful to future researchers when evaluating different decision processes in data sharing initiatives. Additionally, the BPMN format makes it interoperable and relevantly easy to adapt to other solutions.

The decision-making follows this flow:

1. The process initiates when a third party submits a request.
2. The admin can then decide if they will customise the rules that will govern the deliberation0s.
3. When the request deliberation start time passes, members will view the request.
4. Members can comment, vote, and/or submit an offer before the request deliberation ends.
5. When the request deliberation end time passes, the system will calculate the result of the request deliberation.
6. If the members approve of the request, the admin will be asked to approve the policy.
7. If the members don't approve of the request, the system will see if any offers were submitted during the request deliberation period.
8. If none exist, the process ends.
9. If there are offers, the offer deliberation process begins and members will view the offer(s).
10. During the offer deliberation process, members can comment on the offer(s) and/or rank the offer(s) as well as a Reject All option.
11. When the offer deliberation end time passes, the system will calculate the result of the offer deliberation.
12. If the reject all option wins, the process ends.
13. If one of the other offers wins, the third party will have to agree with the new terms.
14. If the third party rejected the offer, the process ends.
15. If the third party approves the offer, the admin will be asked to approve the policy.
16. If the admin rejects a policy, the process ends.
17. If the admin approves a policy, an agreement will be created and access granted to the third party.

The rules in step 2 include when the deliberation starts and ends, as well as the percentage of the members needed to approve a policy. The default time for deliberation is 24 hours after the request is submitted so an admin essentially has 24 hours to decide to customise the rules.

3.7 Summary

This chapter specifies key design choices for the DataConsensus application. These design choices mark the DataConsensus application apart from existing data cooperatives and data sharing initiatives. One of the most obvious differences is that the DataConsensus decision process involves direct democracy instead of representative democracy, which is the case for Midata Coop [67]. Another interesting difference is that the Midata Coop allows non-members to participate in the data sharing [68]. Additionally, the technical framework for the DataConsensus differs from the Salus Coop, which utilises smart contracts and blockchain instead of linked data [53]. Comparing the efficiency and engagement of the different data cooperative frameworks would be an intriguing analysis, but it falls beyond the scope of this project.

Nevertheless, this chapter outlines the key requirements for the DataConsensus application. Specifically, it outlines the defined community, actors, user stories and use cases. Additionally, it presents two key contributions presented in this project: the ODRL Profile for Collective Policies and the BPMN diagram of the decision process for the DataConsensus application. These contributions were core building blocks in developing the application and will hopefully benefit future researchers. An overarching aim in these design choices is to foster collaboration, as evident in the multiple staged decision process and the expanded constraints available in the OCP. This will contribute to the end goal of creating an application that facilitates deliberation of data requests in an collaborative manner.

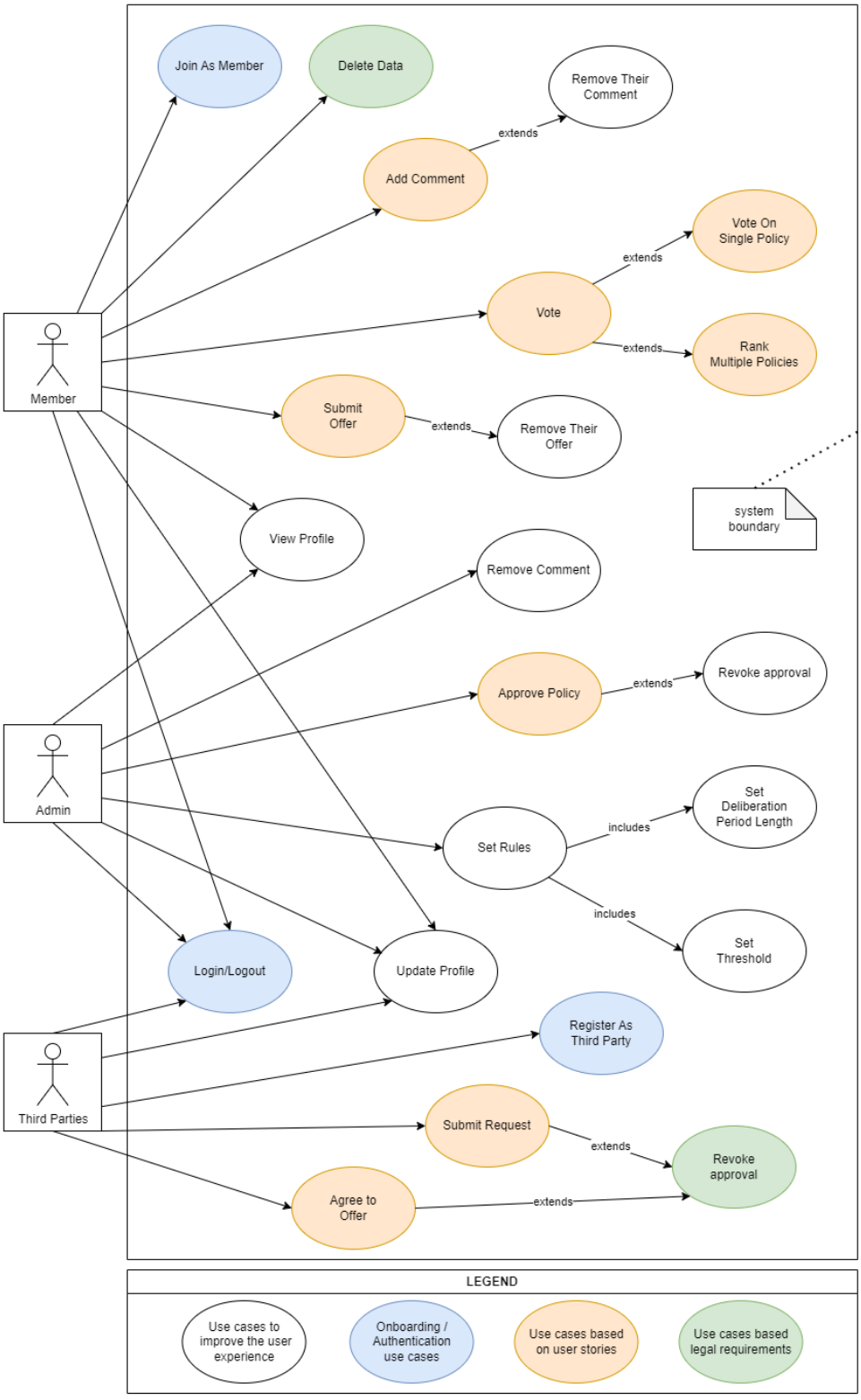


Figure 3.1: The use case diagram for the DataConsensus application, segmenting the use cases by theme

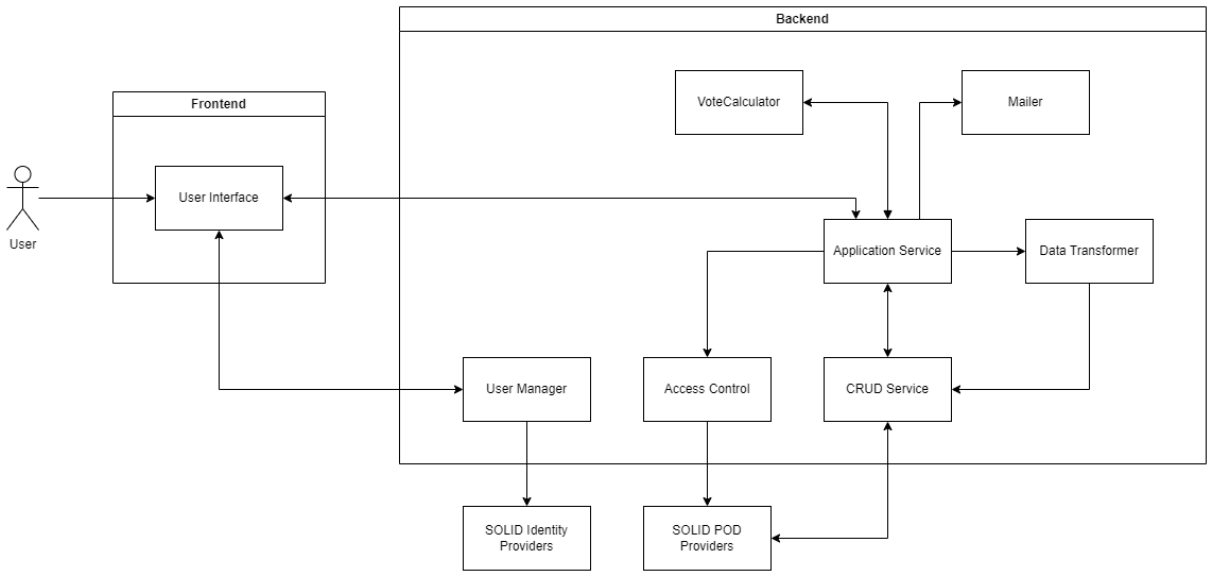


Figure 3.2: Functional Architecture Diagram of the Data Consensus Application

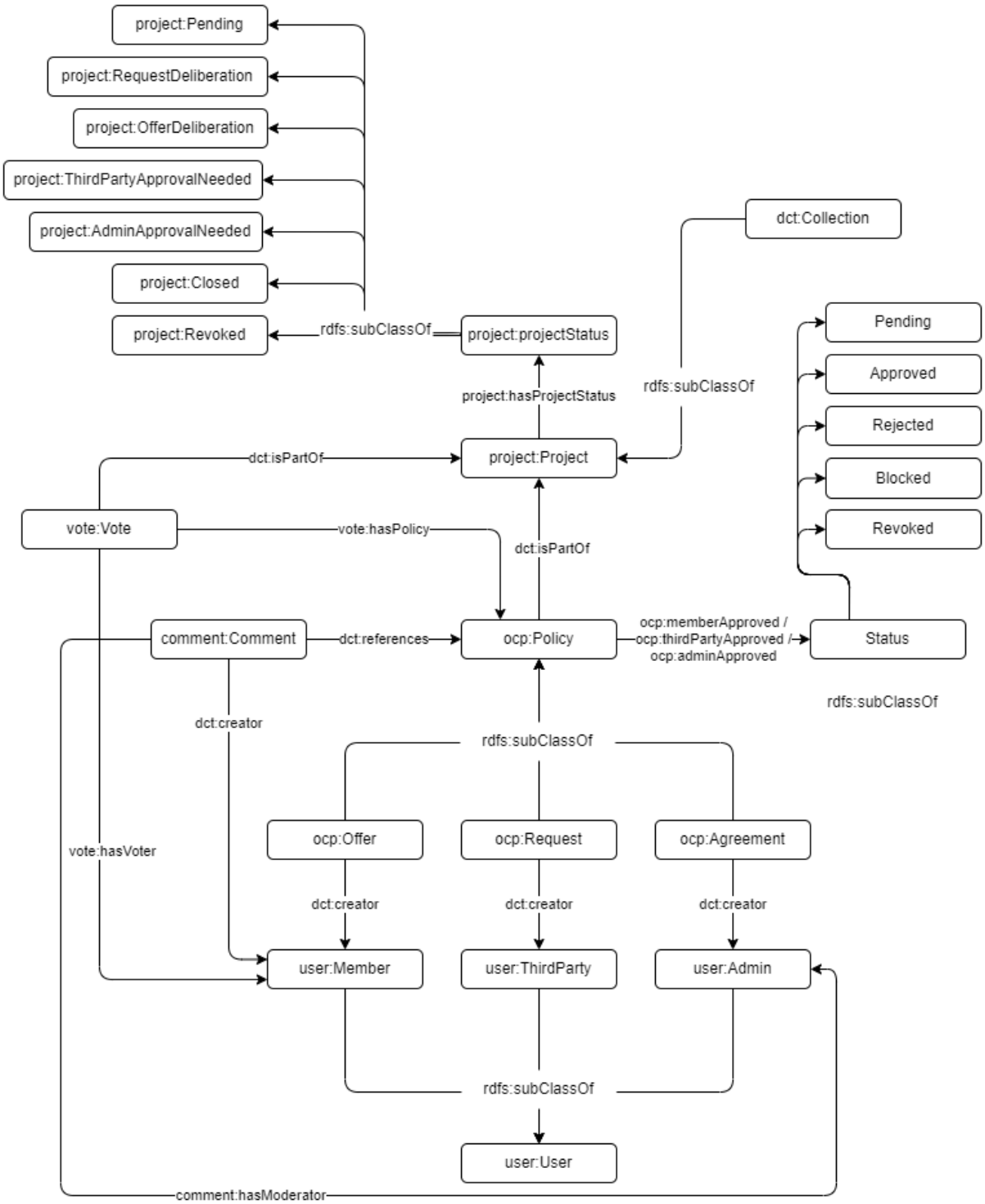


Figure 3.3: Ontology diagram highlighting the relationshps between the classes defined in the DataConsensus schemas.

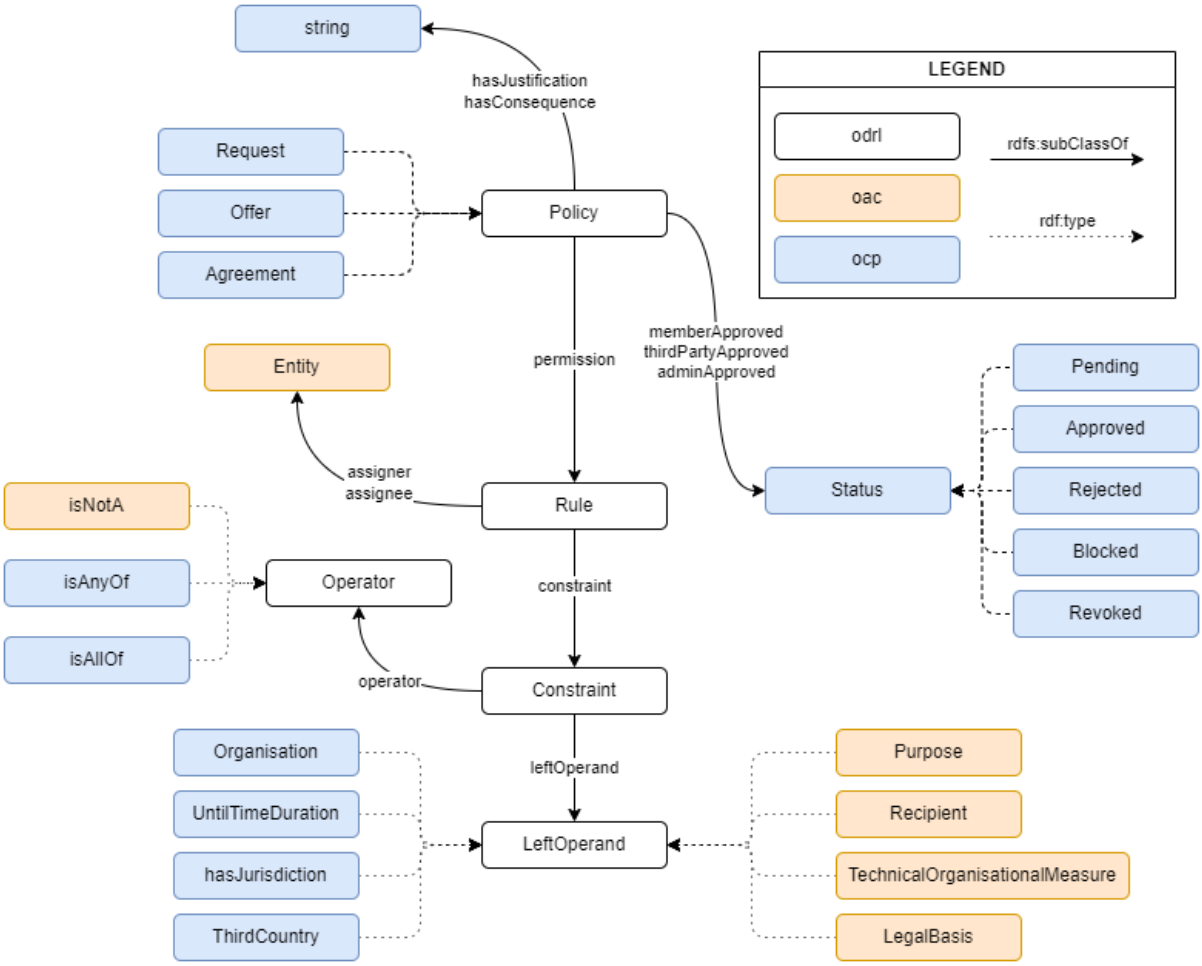


Figure 3.4: Comparison between the OCP, OAC and ODRL.

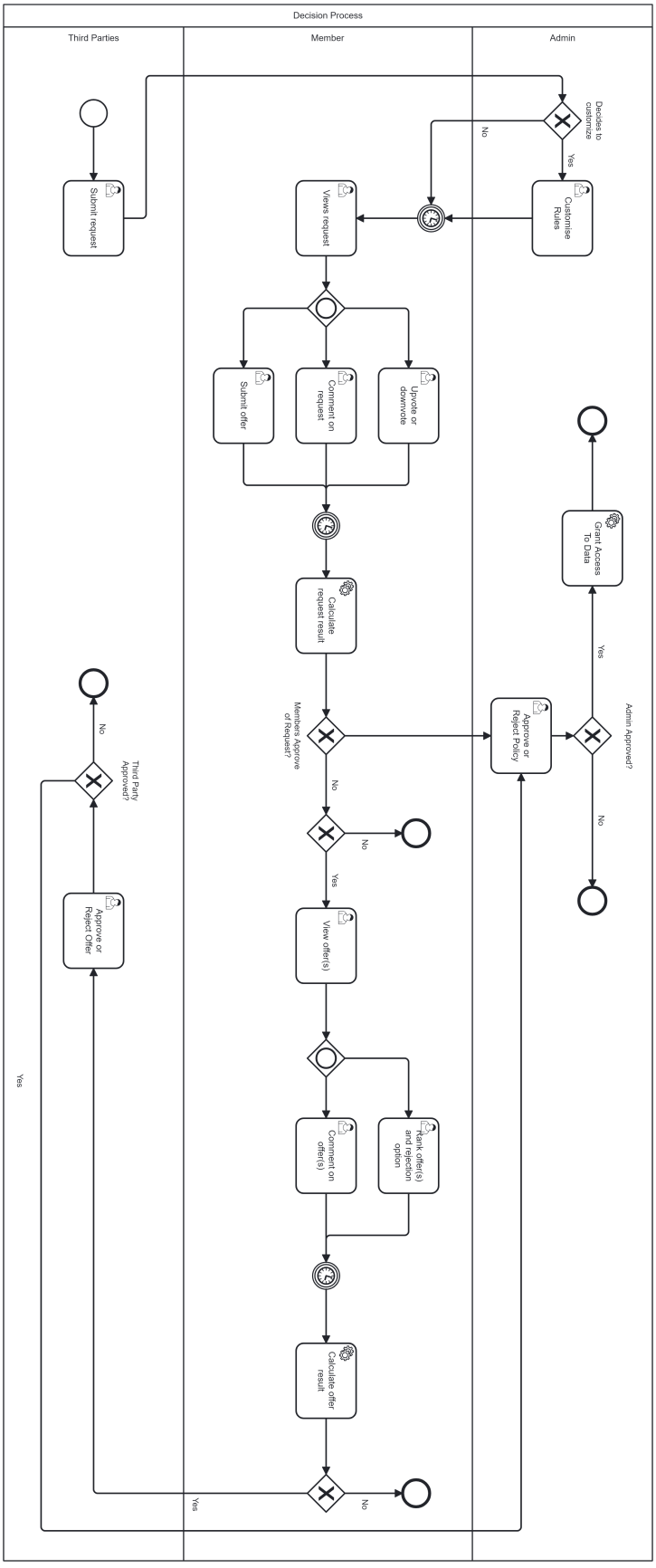


Figure 3.5: The BPMN Diagram for the decision process of the Data Consensus application

Chapter 4

Implementation

This chapter describes the implementation of the design discussed in the preceding chapter. Section 4.1 will outline the technical architecture used. Section 4.2 will explain the development environment and instructions for the set-up. Section 4.1.3 will describe the file structure of the application's solid pod. Section 4.4 will walk through the user interface and user flow of the application. Section 4.5 explains the process of updating the DataConsensus application and how adaptable the system is for future development. Section 4.6 will discuss the implementation issues encountered during the development of the application, followed by a summary in Section 4.7.

Both the code and pod files for this application are available publicly on github at `https://github.com/bod777/DataConsensus.git`.

4.1 Technical Architecture

The application is specifically a web application, built for desktop and was not suitable for mobile users. Figure 4.1 presents a diagram of the technical architecture implemented in this project. There are three layers to the architecture - the UI layer, the application layer and the data layer. Further explanations of the technical architecture layers are given below.

4.1.1 UI Layer

Also known as the frontend, the UI layer is responsible for user interactions. This layer was built using the Angular framework. Additionally, most of the UI components came from the Material Design libraries for Angular. It should be noted that this layer had no interaction with solid and relied on the HttpClient library between it and the application layer to fetch relevant items.

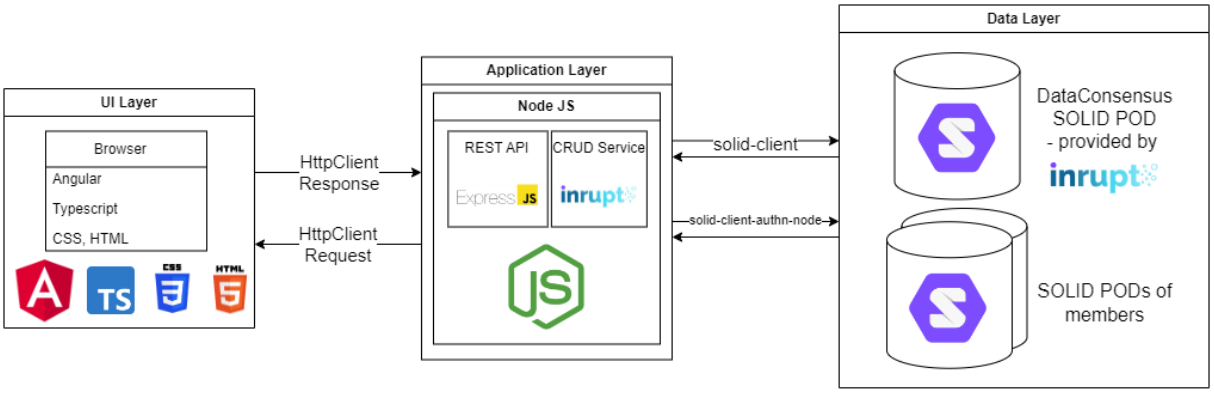


Figure 4.1: Technical Architecture Diagram of the DataConsensus application

4.1.2 Application Layer

Also known as the backend, the application layer serves as the middleman between the frontend and solid. The server is built using Node.js with the Express framework and they were selected as the student had the most experience in these technologies. The application layer was the location of the application's RESTful endpoints. As such, it is responsible for handling requests from the UI layer and transforming the data both to and from the data layer into a suitable format. It is also responsible for the CRUD (Create, Read, Update and Delete) operations, as well as the access control operations.

It communicates with the data layer using the Inrupt's JavaScript libraries @inrupt/solid-client, along with the @inrupt/vocab-common-rdf library for referencing key RDF vocabularies. The other key interaction with solid is the authentication process, which is facilitated using the @inrupt/solid-client-authnnode library. The specific explanation for this authentication flow is described in Section 4.2.

The backend pseudonymised the data pool using the crypto library. It also generates the uids when adding data to the solid pod using the uuid library. Finally, the backend has mailing functionality using the NodeMailer library.

4.1.3 Data Layer

The data layer is entirely built on solid. This was a deliberate choice to experiment and demonstrate the full abilities of solid and linked data in data sharing applications like DataConsensus. An alternative approach would have been choosing a non-RDF supported database, such as Firebase, MySQL or MongoDB, for the vote and comments data. This approach may improve performance and scalability and exploring these options could be potential future work for research in this area.

The most important aspect of the Data Layer is that the application not only accesses its own solid pod, which is where the policies, comments and other data is stored, also interacts with the solid pods of the application's members. DataConsensus only fetches the given data source from a member's pod and it can only do this within 5 minutes of authentication. In the current iteration, DataConsensus only fetches that resource when the member first registers or when they update their Data Source information on their profile page, and they successfully authenticate their WebID. This means that the number of times the application accesses a member's pod is initially controlled by the member.

Pod File Structure

In this section, the file organisation of the DataConsensus solid pod is explained. This section should aid anyone expanding on the application. Appropriate pod files for a demo of the DataConsensus can be found in the Pod directory of the GitHub repo mentioned at the beginning of this chapter.

Figure 4.2 presents the file structure inside the DataConsensus pod. The app folder is in the base directory of the solid and all the turtle files that store the data related to the app such as the policies and users are contained there. The pseudonymised pooled data is stored datapool.csv in the pool folder. The schemas folder contains all the RDF schemas related to this project, including the ODRL profile for collective policies (OCP).

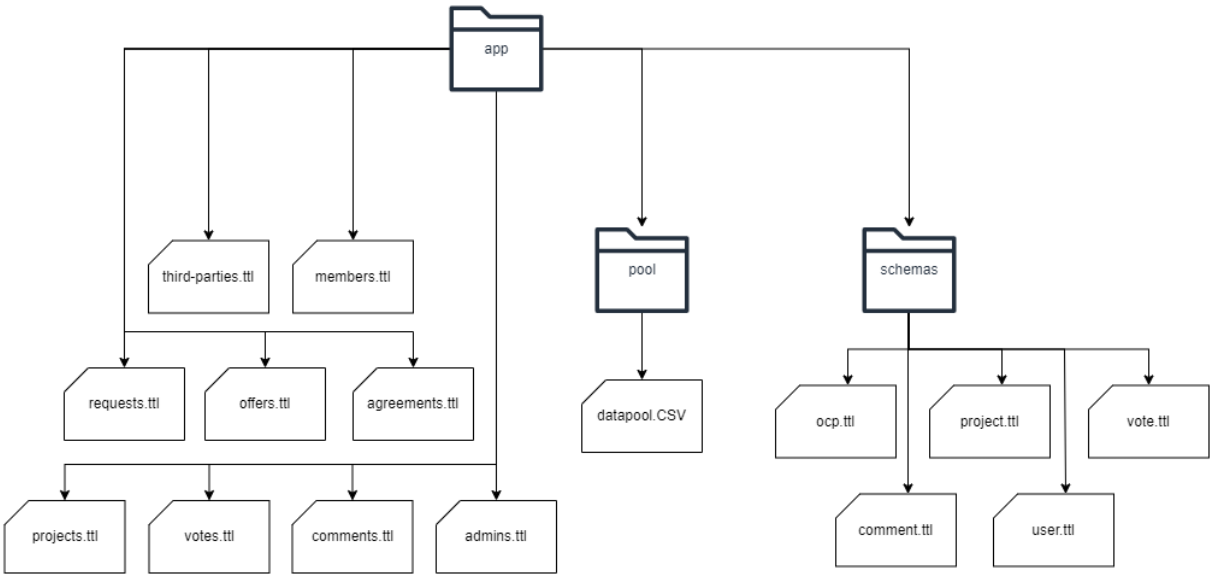


Figure 4.2: The file structure in the DataConsensus pod, including schemas and the pooled data

The main data files are only accessible by an entity logged in as the pod's webID. The Resource.csv is accessible to the approved third parties and the pod's webID. The

schemas folder is publicly accessible and can be found at: NEED THE URL.

4.2 Authentication

As explained in Chapter 2, the solid protocol manages the authentication process itself. There are two authentication processes within the DataConsensus application: the app authentication and user authentication. The user authentication occurs at registration, login and upon updating the data source as a member and it authenticates the user and allows the application to access the user's pod. The implementation of the user authentication mostly follows the instructions found on Inrupt's documentation on Authenticate (Node.js Web Server) [3]. When a user begins the user authentication process, the frontend redirects to the relevant authentication endpoint in the backend. This endpoint brings the user through the solid authentication process and redirects back to the backend, which then redirects to the frontend, with the sessionID, webID or relevant data stored as URL parameters. These parameters are saved in the local storage and if the user logs out, these items are removed.

The app authentication allows the application to access the application's own pod. It follows the Inrupt's instructions for Authenticate with Statically Registered Client Credentials on Authenticate (Node.js: Single-User App) [2]. Since the solid session ends after 5 minutes and this is unchangeable, the app authentication process runs on an interval to ensure the application can always access the pod. The alternative approach was to use refresh tokens, as described in inrupt's documentation. However, this approach appears to be deprecated and to initially obtain a refresh token was too difficult. This appears to be a significant oversight with regards to solid and its suitability for many applications. Therefore, while solid worked well for user authentication and fetching a user's resources, other linked data storage systems could be more appropriate for future iterations of the DataConsensus application.

4.3 Development Environment

The IDE used for this project was VSCode, with a Windows 11 Home 64-bit operating system. Additionally, the web browser used during development was Google Chrome. Additionally, Postman was used to test the endpoints of the backend. Table 4.1 presents the versions used for the environments and frameworks:

Inrupt's PodSpaces and PodBrowser were the primary tools when setting up and managing the pods required for development [4]. The Inrupt's PodSpaces pods is hosted by

Framework/Environment	**Version**
Angular	16.1.0
Node.js	18.13.0
Express.js	4.16.1

Table 4.1: Versions of the development environments and frameworks used in this project

Amazon and hosted in Germany [96]. PodSpaces' OpenID Connect link is https://login.inrupt.com. A pod is linked to a webID and to verify an account in Inrupt's PodSpaces, an email address is required. Therefore, a number of gmail addresses were created along with the webIDs to reflect the different actors involved. See Appendix C.2 for the accounts created for testing in this project. Furthermore, instructions for setting up the application can be found in Appendix C.3 and additional guides to the software development can be found in Appendix C.4

4.4 User Interface

In this section, the key pages and components of the DataConsensus user interface will be examined.

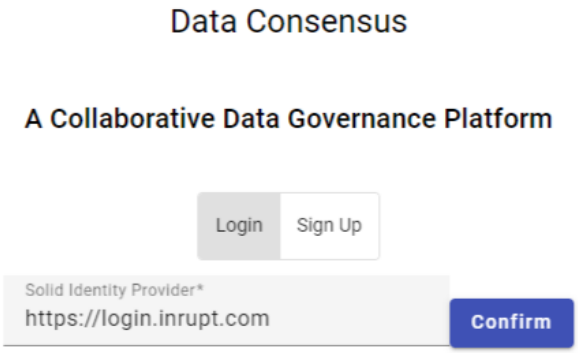
4.4.1 Logging In

The logging in interface is relevantly straightforward and was inspired by the design in the solid tool proposed [91]. Figure 4.3a and Figure 4.3b presents two different views of the login page. Figure 4.3a details the process, requiring the user to input their solid identity provider, defaulting to `https://login.inrupt.com`, associated with the test accounts. Upon clicking "Confirm", the user will be sent to the external Solid authentication process. This is followed by a validation of their webID against DataConsensus' records to ascertain their existence and retrieve their specific user type.

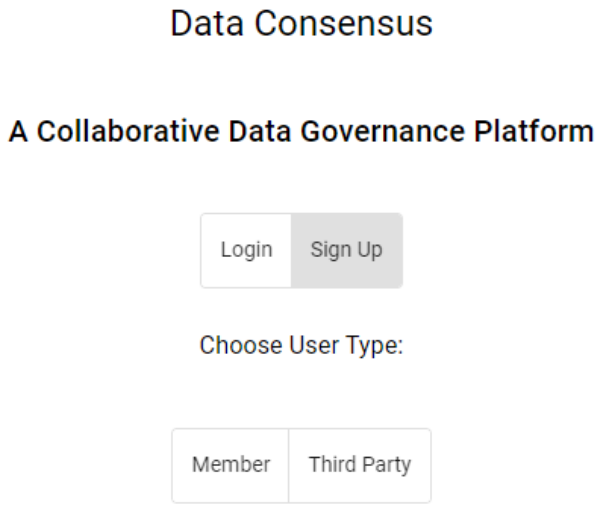
Figure 4.3b illustrates the page when a user opts to sign up instead of login. After selecting a user type, the user is guided to the registration page, which will be further described in Section 4.4.2.

4.4.2 Registration

Two registration pages exist: one for members and one for third parties. Both are structured as forms, requesting a name, email, and solid pod provider. Users joining as third-parties must also specify their organisation type and provide a description.



(a) Image of Login Page, including the Solid Pod Provider input



(b) Image of Login Page, including the view for when the SignUp button is selected

The /member-signup page will additionally require the URL of the user's entries.csv, an export from NightScout software, to. This URL, referred to within the system as the Data Source, is used to locate the user's data in their pod for the shared data pool. Figure 4.4 depicts the /member-signup page. An important aspect of the page is the note prior to the confirm button, clarifying how the application will access their data to the joining member.

The initial information collected during registration could be broadened to include members' date of births or genders, or additional contact details and research information for third parties. However, a more limited scope was chosen, as this application serves as a proof of concept.

Upon clicking "Confirm" on the registration pages, users are directed to the solid authentication process, mirroring the login procedure. They will return to the registration page, where a loading bar and notification alerts will signal the ongoing registration. Once registered and logged in, users are redirected to the appropriate home page.

4.4.3 Home

Once logged in, a vertical menu will appear at the top of every page, with buttons varying by user type. All users see Home, Profile and Logout buttons. Members and admins also see an Agreements button, while third parties see a Submit Request button.

As mentioned previously, the homepage is where users are redirected when they log

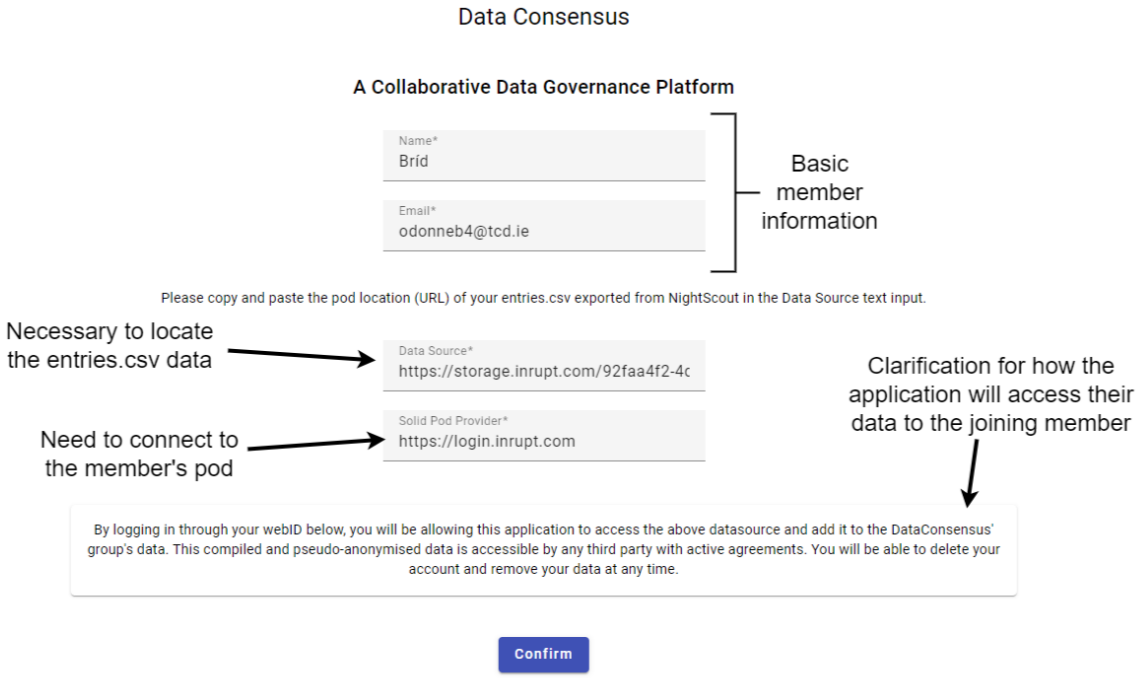


Figure 4.4: A image of the member sign up page.

in. There are different homepages for each actor, however it should be noted that the only difference between the admin and member homepage is the title. Figure 4.5 shows this homepage. There is one key feature on the home page - a table of the projects in the system. Users will be able to click on the requester and navigate to their profile page or click on the "View" button to visit the project's page.

The third party home is also focused on a table of projects but only the projects that the third party is involved in. Another difference is that there is a secondary status for whether there is an in-date agreement.

4.4.4 Project Page

Project page is the main tool for viewing and interacting with a project and its related policies. The project page has a sidebar for navigating to different policies of the project, along with an Overview of the project. If there are existing offers linked to a project, an Offers tab will appear at the sidebar, similarly if there is an agreement, an Agreement tab will also appear. Figure Figure 4.6 shows the Overview tab of a project, including a description of the research project and other key information. The notable features of this Overview is the Do you approve feature, this appears for an admin user when the project status is Admin Approval Needed or a third party user when the project status is Third Party Approval Needed.

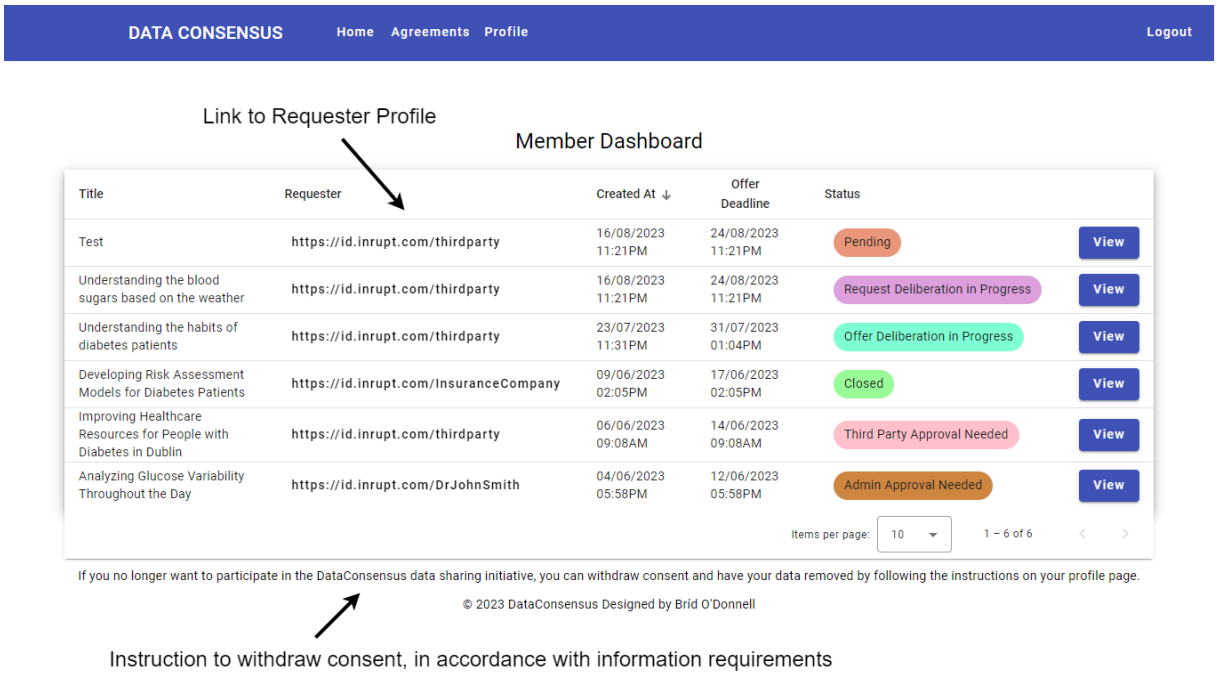


Figure 4.5: Image of the Member Home Page

There are other conditional features of the Overview drawer. When an admin is viewing a "Pending" project, they will be able to see a "Change the rules for this project" form, which allows the admin to change the start and end times of the deliberation, as well as the threshold for passing a policy. Furthermore, if a third party has access to the pool data, instructions to access that data is seen on the overview for them.

Figure 4.7 presents the Request tab for a request that has been rejected by the members. The key features include the policy statuses for the different actors, the constraints section and the comments section. The policy statuses illustrate the stage the different policies are at while the constraints section simplistically and concisely present key information from the policy to the members. Furthermore, the comments section is a key deliberation tool for the members, allowing them to share opinions and ideas which can encourage collaboration and easily track sentiments among the members. Note that the perspective in Figure 4.7 is an admin's, therefore the "moderate" buttons are available to ensure the community is safe and friendly in its discussions.

Figure 4.8 shows the upvote and downvote buttons which are visible to members during the Request Deliberation. Below these buttons is a button that brings the member to the Offer Builder page.

The upper half of the Offers tab during the Offer Deliberation can be seen in Figure 4.9. The difference between the Request and Offers tab is the ranking feature, which is only viewable to members during this period and below this are horizontal tabs of the

different offers submitted. A constraints section and a comments section are featured beneath each individual offer, allowing users to view and compare each offer.

It should be noted that if viewing a policy as its creator, the user will have the option to delete the policy so long as the policy has not been deliberated on yet. As for an agreement, any admin can remove access to the data pool for the third party. This option was given in case a third party broke a condition of the agreement.

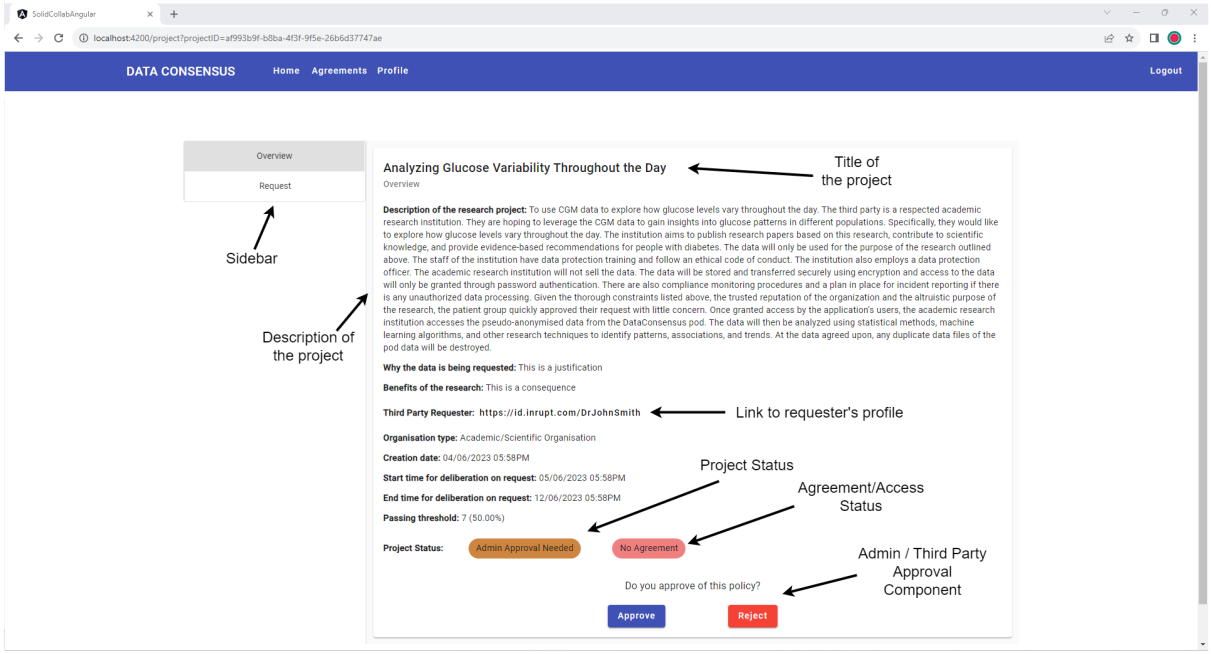


Figure 4.6: Image of the Overview tab of the Project Page. This is the Project Page for the Academic Researcher user story and it is currently being viewed by an Admin.

4.4.5 Policy Builders

There are two policy builders, one for third parties to create their requests and one for the members to create an offer in response to a request. Conflicts within the policies, with constraints and rules contradicting each other, should be avoided because of the design of the OCP policies and closed inputs of the policy builders.

The Request Builder page can be seen in Figure 4.10. Aside from the first four inputs, the rest all translate into constraints for the ODRL policy. The builder tries to utilise dropdowns as much as opposed to making the form streamlined. The textarea inputs do increase the size and effort of the builder but they provide key information that can aid the member's deliberation and most are required based on legal and ethical compliance. These requirements will be discussed in further depth in Chapter 5. The final two inputs

are not required and should only be filled when the data will be shared abroad.

The Offer Builder page is similar to the Request Builder, however it excludes a number of textarea inputs that describe the research. This is because a member should not be in the position to dictate the research at such a high level. Instead they can make adjustments to an existing request, such as requiring more technical/organisation measures or removing permissions to sell data or insights etc.

4.4.6 Profile

The profile page's appearance varies based on user type and whether a user is viewing their own profile. In the former's case, the user will be able to update their profile information and in the latters, it can be used to reach out privately to other members, third parties or admins since a user's email will be available. Nearly all webIDs seen in the application are buttons linking to the corresponding profile. Please note that third parties are not able to view the profiles of members to ensure their privacy.

The Member Profile, seen from the owner's perspective in Figure 4.11, is the most complex. The member can update their name, email and data source. The data source update is separate from the others because when the "Update Data" button is clicked, the system will replace the member's data in the pool with the data in the new data source URL. This will require the member to log in again through the solid authentication process to ensure that the system can access the member's pod.

Finally, the member can delete their user data. This will remove the member from the member list, remove their data from the data pool and remove their WebID as a data subject in the existing agreement policies. The system will also send an email to all agents who currently have access to the data pool to inform them that their data is out-of-date and they must delete any existing downloaded files of the data and replace it with the new data in the pod. This ensures GDPR compliance.

4.5 Design Updates

After the initial development of DataConsensus and a number of successful demos, a number of missing legal requirements were missing regarding the request information. This information required included detailing the benefits and risks of the research we performed, the justification for requesting this specific data and whether or not the data will be shared outside of the Irish state. To ensure that the application fulfilled as many legal requirements as possible, an update to the application occurred.

Firstly, suitable vocabulary was identified to fulfil these information requirements,

such as dpv:hasConsequence, dpv:hasJustification and dpv:ThirdCountries. The existing policies in the pod had to be updated to reflect these changes. The update to frontend involved UI changes to the submitRequest, submitOffer and constraints components, and it mostly involved replicating existing code and creating new variables. The update to the backend was a bit more complex, mainly updating the policy class and createPolicy function in the Policy Service. This process took one evening, significantly less time than the initial development. More modularity could be achieved in the software and would be necessary to offer customizability for the data communities using the application. Adding the customizability should be considered future work to improve the application. Regardless, this update shows that the application's components can be modified independently without causing failures in other areas.

4.6 Implementation Issues

One of the significant challenges encountered during the implementation phase of this research was the underdeveloped community and documentation surrounding SOLID technology. The community of developers, researchers, and enthusiasts focused on SOLID appears to be relatively small. This scarcity created a tangible obstacle in locating existing projects, specific problems, or most importantly solutions that bore resemblance to the particular challenges faced in the context of this project.

Solid has been released since 2016, however it hasn't the same adoption as other decentralised projects such as ActivityPub which was released in 2018 and will soon be supported by Tumblr and Threads [83]. It's unclear why solid has had such slow growth, but it could be due to the steep learning curve of linked data. Another reason for its lack of popularity could be a few key technical limitations posed by Solid.

One substantial limitation in Solid technology is the lack of server-side support for SPARQL, a query language for databases, despite Solid's inherent support for Linked Data [88]. The absence of server-side SPARQL support curtails the ability to perform single queries as efficiently and elegantly as would be possible with traditional database systems. There are community suggested libraries to enable client-side SPARQL querying [88] [57] [59], but implementations of these libraries failed. Disappointingly, there are no plans to implement server-side SPARQL in the solid specification [74].

While it was indeed possible to access the necessary data and extract relevant information, the process required alternative methods that might be seen as less efficient or intuitive compared to standard database querying. Furthermore, the lack of SPARQL support could potentially introduce scalability issues as the application grows. A substantial increase in policies and members might stress the current data retrieval methods,

leading to performance bottlenecks or maintenance challenges. Further research and development may be required to understand the full extent of these potential limitations and to devise strategies for mitigating them, especially if Solid technology continues to be an area of interest for large-scale applications.

One final limitation to Solid that has previously been mentioned in Section 4.2 is the app authentication. Allowing applications to easily access their own or other pods is a challenge that is still open within the Solid community [86]. The method implemented by this project is inelegant and is fault-prone. Without an easy way to authenticate an application to access its own pod, it's hard to imagine applications being built solely on Solid and that will be a major deterrent to the community growing more.

4.7 Summary

This section examines the technical architecture, justifying its choices and the responsibilities of each layer. The two authentication processes utilised in the DataConsensus application are explained and key information related to the development environment and set-up are given. A description of the user flow and the user interface is provided in Section 4.4. Furthermore, Section 4.5 explores updating the existing code, illustrating the application's suitability for future development. This is followed by a discussion of the implementation issues, which largely focuses on Solid's technical limitations. This section endeavours to provide brief explanations to the key software features and user interface of the application, while highlighting any unusual or complex features that might perplex future developers building off of this work.

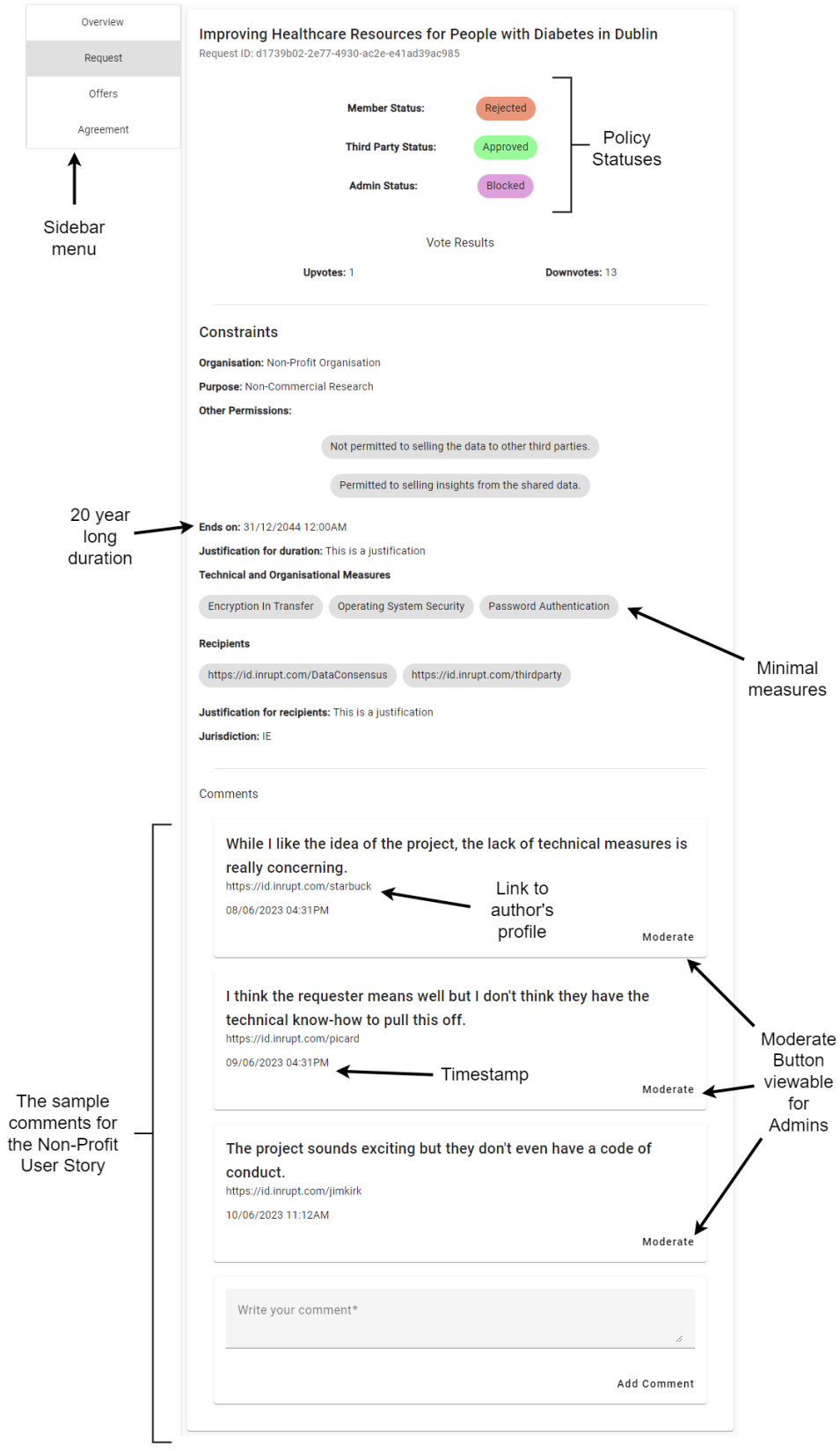


Figure 4.7: Image of the Request tab of the Project Page. This is the Project Page for the Non-Profit Organisation user story and the request has been rejected by members after a deliberation and discussion in the comments.

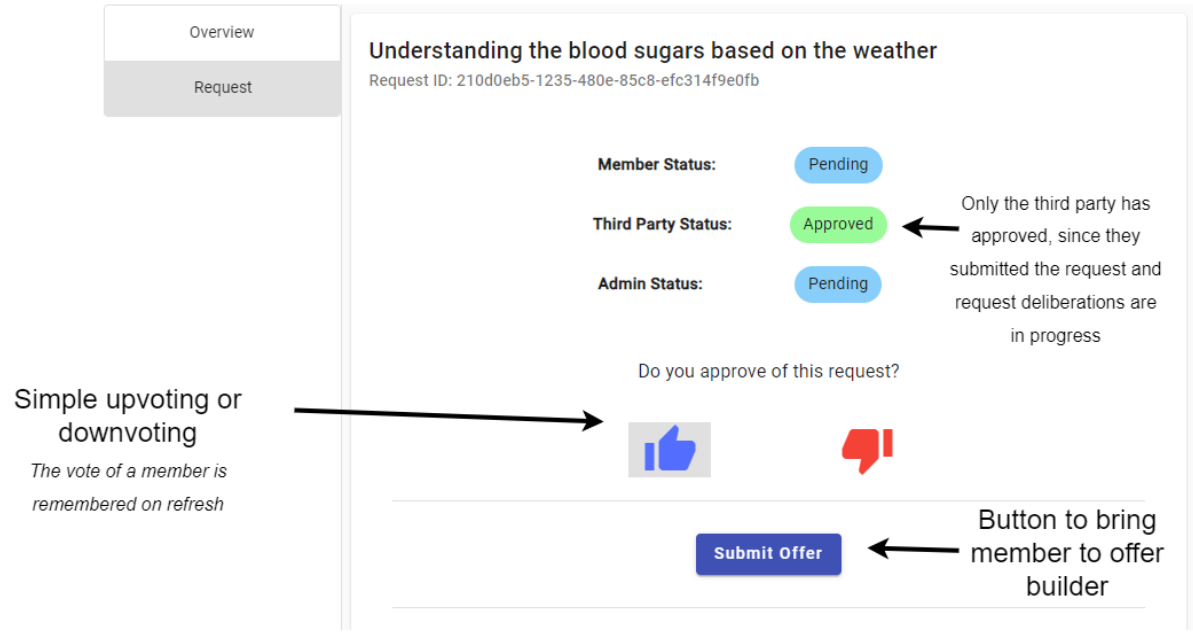


Figure 4.8: Image of the Request tab of the Project Page during the request deliberation.

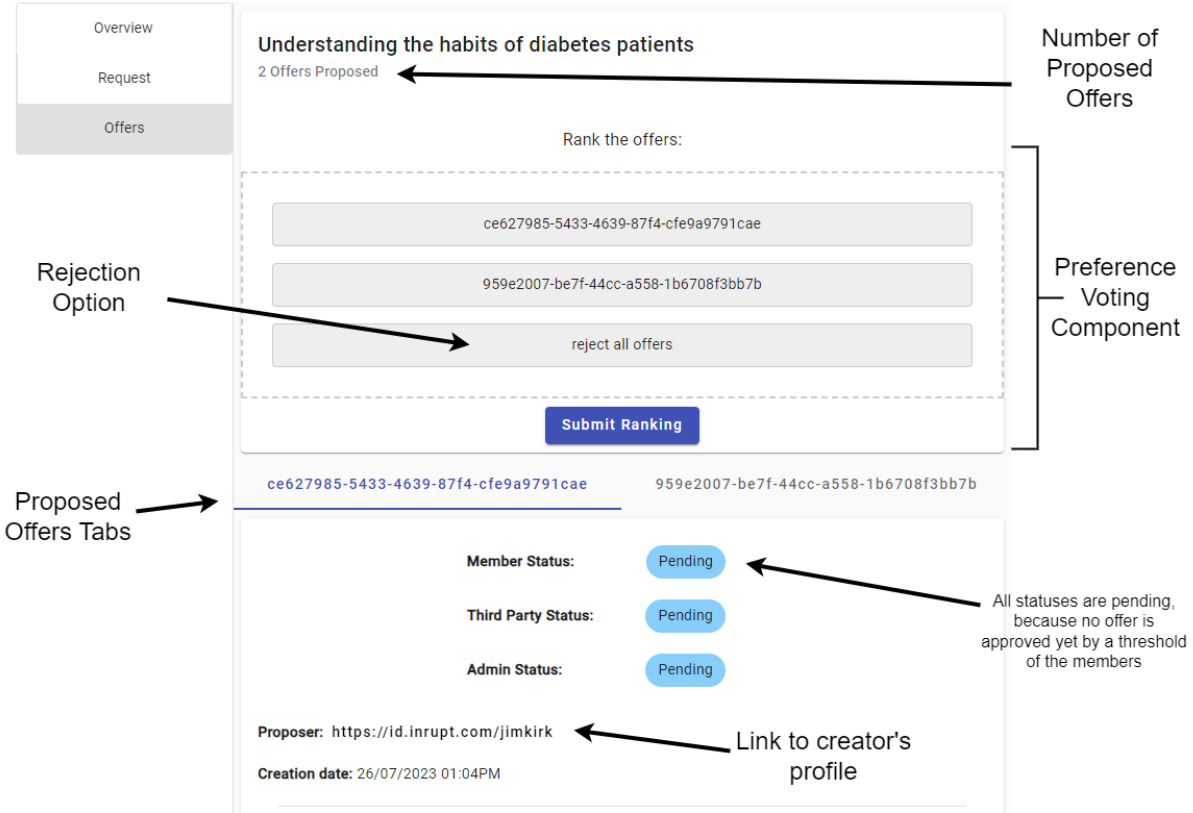


Figure 4.9: Image of the Offers tab of the Project Page during the offers deliberation.

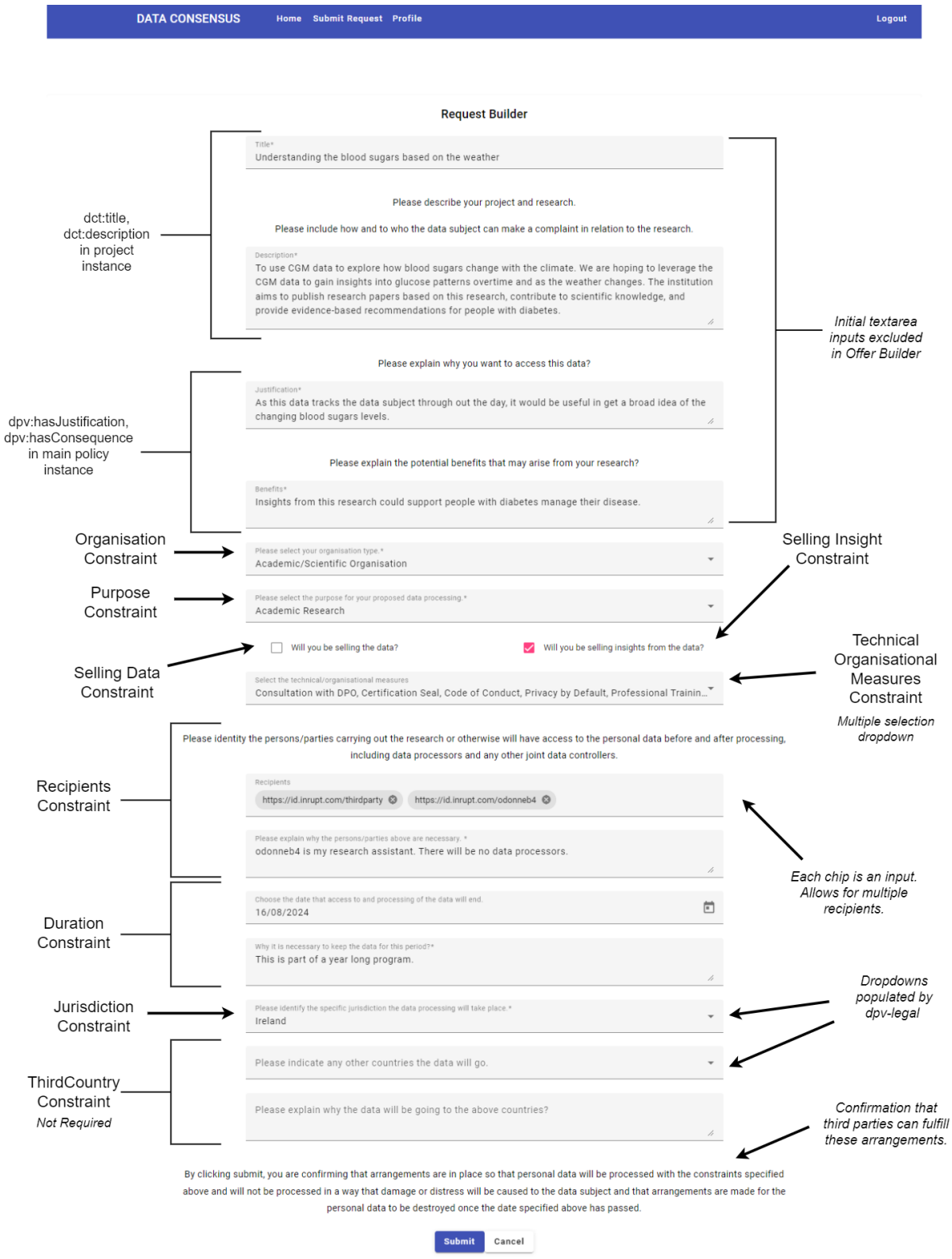


Figure 4.10: Image of the Request Builder

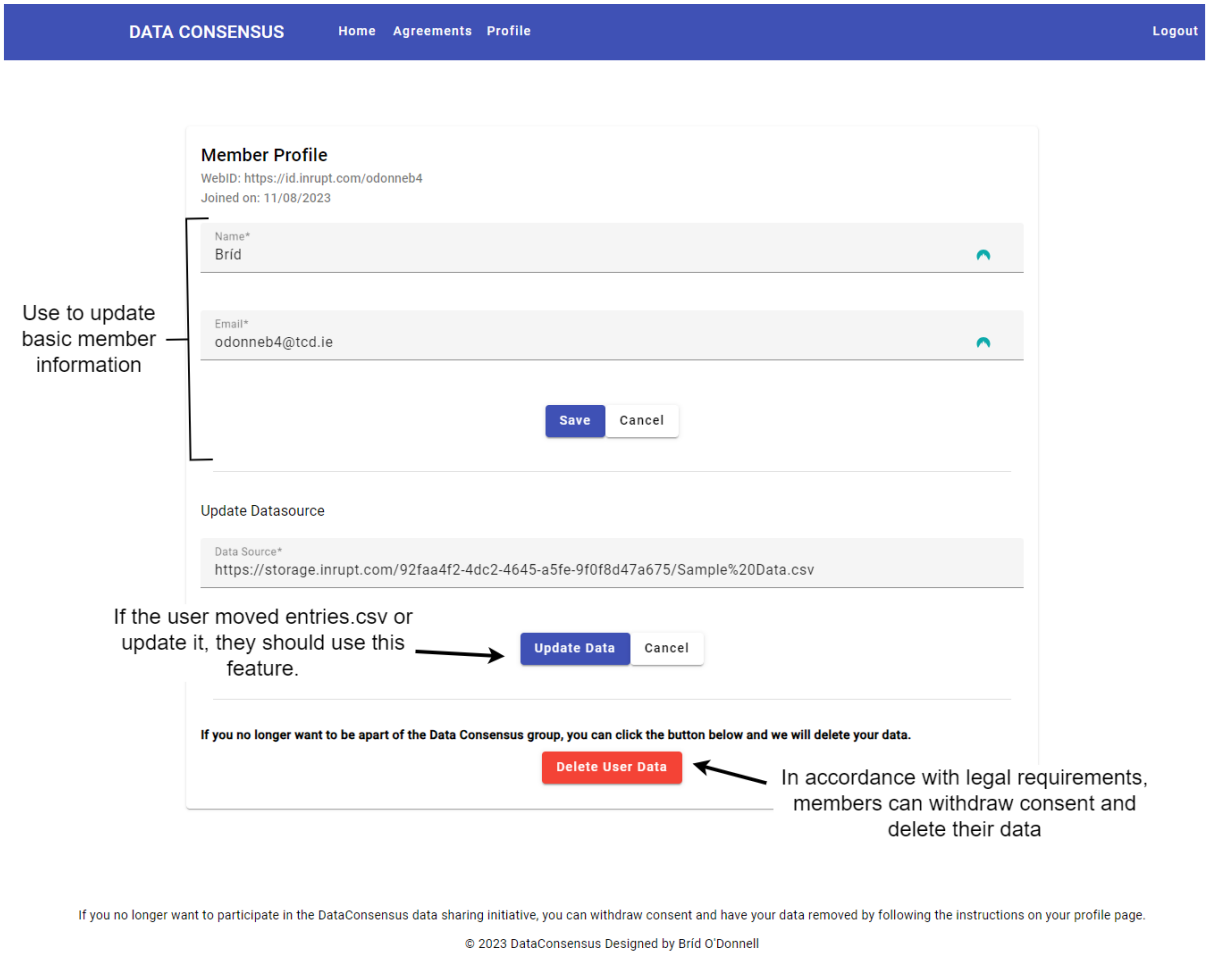


Figure 4.11: Image of a Members' Profile from the perspective of the person who owns the profile

Chapter 5

Evaluation

In this chapter, the DataConsensus application will be evaluated against its functional requirements. This chapter will also propose the Data Policy checklist, which outlines the criteria when obtaining consent for health data sharing. Following this, Section 5.2 presents a set of guidelines for building a fair and sustainable data commons. Finally, this chapter will conclude by evaluating the security and privacy of the Data Consensus application.

The data policy checklist and community-related guidelines will be used to evaluate the current iteration of the DataConsensus application. Both are contributions to the field of collective data governance but serve different purposes. The checklist covers the legal requirements and best practices for obtaining consent, specifically for data commons sharing health data with researchers within Ireland. The guidelines serve a broader audience, including researchers or operators of data commons. Given the ethical, regulatory and professional standards surrounding consent, the checklist is separate from the guidelines and its criteria should be more strictly followed than the guidelines.

The functional requirements of the application are outlined in Chapter 3. The most important requirements are those specified by the user stories. These requirements are achieved by the application in this current form.

5.1 Data Policy Checklist

The data policy checklist covers 32 criteria for the information that must be shared with data subjects when obtaining consent, along with 6 criteria related to the overall procedure for obtaining consent and the processing of the data. It was formulated based on four different sources. The first source was GDPR, predominantly the data processing rules found in Article 5 and the conditions for consent found in Article 7. These are the most

basic legal requirements that must be followed when asking for consents and are more straightforward to assess compared to other items found in the checklists. Following this, requirements based on the Irish regulatory regime for health research were identified and included.

In the 2021 amendments to the Health Research Regulations (HRR), researchers were required to comply with "international best practice on the ethical conduct of health research" when obtaining consent [28]. International best practices can be a vague requirement to access requests against. Fortunately, documents from the Department of Health and other government organisations do provide some guidance, though they maintain that the guidance is not exhaustive. In cases where researchers do not include certain information when obtaining consent, they must be able to justify why, if asked by the data subject or the Data Protection Commission [28]. Therefore, the checklist represents both legal requirements under GDPR and HRR and best practices that should be followed in accordance with HRR.

Of the two Irish government documents used in this checklist, [28] provides the most comprehensive information criteria, with clearly listed items to include. However, [28] was published in 2018 following the initial adoption of HRR so [29] which was published following the 2021 HRR amendment to ensure the checklist was up-to-date with Irish regulations. Finally, the template consent statement published by Salus Coop was consulted. This final source provided a practical example for communicating consent-related information in a health research scenario.

There were no contradicting criteria listed in any of the four sources. Generally, all sources required that when asking for consent, the purpose be specified, withdrawal be allowed and the information be provided in a clear and accessible manner. This would be following the legal requirements outlined by GDPR. Generally, [28] is the most comprehensive criteria and even the later [29] does go into as much detail regarding information on the data recipients, DPO and other persons involved in the research. The [89] is interesting, since Salus Coop isn't an Irish body, it doesn't have to follow the HRR but it does cover much of the information required by [28].

Table 5.1 presents the criteria items in the data policy checklist, along with its sources. The checklist contributes to the research area, as the information regarding legal requirements and best practices for requesting consent are disbursed and confused. Additionally, they are generally not formatted in a way for easy evaluation, especially against software systems.

The evaluation of the DataConsensus application against the data policy checklist can be seen in Table 5.2, along with the explanation for how the criteria was fulfilled or not. 18 criteria were fully fulfilled. 12 requirements are only partially fulfilled and 7 requirements

are not covered by the DataConsensus application. Finally there is 1 requirement that is not applicable to the DataConsensus application. As seen in the explanations for the partial fulfillments, this project held itself to a high standard for full compliance and future work should endeavour to achieve full compliance with the criteria.

#	**Criteria**	**Source**
	Information that should appear in the policy	
1	Title of the research project	[28], [89]
2	Purpose / scope of the research project	[28], [29],[39],[89]
3	Description of the data required and the justification for processing this data	[28],[89]
4	Description of the potential benefits that may arise from the research	[28], [89]
5	Legal basis for requesting the data	[28], [39], [89]
6	Identification of the data controller or joint data controllerst	[39],[89]
7	Contact details of the data controller or joint data controllers	[28],[89]
8	Contact details of the data protection officer (DPO) associated with the research	[28]
9	Identification of the Principal Investigator (PI) and relationship to the data controller if they are not the data controller	[28]
10	Any person who will be a recipient of the data and why it's necessary for them to access the data	[28]
11	Identification and contract details of any data processors	[28]
12	Justification for why it is necessary to have a data processor	[28]
13	Confirmation that anyone with access to the personal data is bound by either a professional code of secrecy or a contractual code of secrecy or some other arrangement that emphasises confidentiality.	[28]
14	Confirmation that data protection training has been provided to those individuals involved in carrying out research.	[28]
15	Details about anyone who funds or supports the project, including their potential access to the personal data gathered	[28]
16	Description of how and from who the data will be collected from (such as healthcare providers possessing health records)	[28]
17	Length of time the personal data will be kept	[28]
18	Justification for why the data needed to be kept for that length of time	[28]

Continuation of Table 5.1		
#	**Criteria**	**Source**
19	Confirmation that arrangements are in place to destroy the data when this time has passed.	[28]
20	Statement to confirm whether the results of the research will be used or disclosed for commercial purposes.	[28]
21	Description of the data security arrangements in place	[28]
22	Confirmation that a data protection impact assessment or an assessment of the data protection implications of the research has been carried out.	[28]
23	The results of the aforementioned assessments, including any risks and what are the measures taken to mitigate the exposure.	[28], [89]
24	Statement on whether the data leave the State and if so what countries it will go to and why it is going to those countries?	[28]
25	How and to who the data subject can make a complaint in relation to the research	[28], [89]
26	Contact details of the Research Ethic Committee (REC) that gave approval to the research	[28], [89]
27	The date the approval was given by the REC	[28]
28	Reporting arrangements agreed with the REC	[28]
29	Any conditions attached to the research by the REC	[28]
30	Explanation for how consent can be withdrawn and the effect of any such withdrawal	[28]
31	Clarification on whether the individual providing the personal data will be advised of any outcome from the research that would impact directly or indirectly on his or her health	[28]
32	Confirmation that arrangements are in place so that personal data will be processed only as is necessary to achieve the objective of the health research and will not be processed in a way that damage or distress will be caused to the data subject	[28]
Criteria related to the overall procedure for obtaining consent and the processing of the data		
33	Consent is freely given and voluntary.	[28]
34	The request information is given in an intelligible and easily accessible form, using clear and plain language.	[29]
35	Choosing to not grant consent will have no adverse consequences on the individual's treatment.	[28]
36	The data subject easily withdraw consent in a convenient way.	[29], [39]

Continuation of Table 5.1		
#	**Criteria**	**Source**
37	The consent is documented by the controller in written, electronic or other format with a copy of the record of consent provided to the individual	[29]
38	The consent request gives choices to individuals in terms of the areas of research that they want their information to be used in and third parties that they are willing to have their information shared or not shared with	[29]
End of Table 5.1		

Table 5.1: Data Policy Checklist, along the sources for each criteria.

#	**Fulfilled?**	**Explanation**
1	Yes	As seen in Section 4.4.5, the title of the research is clearly stated on the home page and project page.
2	Yes	This is fulfilled in two ways, the first is in the Purpose constraint of the OCP policy where either academic, commercial and non-commercial research is allowed as an input. Furthermore, the description of the research is required to create a project and this can be seen clearly on the Overview section of the project page.
3	Yes	This is fulfilled by the `dpv:hasJustification` property in the OCP.
4	Yes	This is fulfilled by the `dpv:hasConsequence` property in the OCP.
5	Yes	All policies within this system fall under the legal basis of consent in this work, as required by HRR and this is clearly stated by the `odrl:hasLegalBasis` property of the OCP agreement policies.
6	Partial	This is only partially fulfilled as the system currently doesn't allow for two third parties to make a joint request or for a third party to specify another data controller on top of themselves. However, the identification of the third party as a single data controller is clearly specified in the OCP agreement policies.
7	Partial	This is only partially fulfilled for the same reasons above. Additionally, there are limited contact details within the DataConsensus system for the third parties, as only their email is required. This could be expanded to phone numbers or addresses too.
8	No	The system currently doesn't require the contact details of the DPO.
9	Partial	The system currently doesn't have an input or property to specify the PI but their recipient's input does allow the user to specify a PI and justify their inclusion.

Continuation of Table 5.2		
#	**Fulfilled?**	**Explanation**
10	Yes	The system allows this with the Recipients constraint and furthermore augments this constraint with a `dpv:hasJustification` property to explain why the recipients are included.
11	Partial	The system allows this with the Recipients constraint, though it is grouped in with other recipients and this might cause confusion. Additionally, there is no way to provide the contact details for the data processors.
12	Partial	The `dpv:hasJustification` property in the Recipients constraint partially fulfil this requirement but it is grouped with the other recipients which might cause confusion.
13	Partial	This requirement can be fulfilled with the Technical Organisation Measures constraint. However, it could be argued that these specific measures should be separate constraints to avoid all ambiguity.
14	Partial	This requirement can be fulfilled with the Technical Organisation Measures constraint. However, it could be argued that these specific measures should be separate constraints to avoid all ambiguity.
15	Partial	This requirement is partially fulfilled as if the funder or supporter has access to the data, they should be identified in the Recipients constraint but lack of its own constraint might lead to ambiguity.
16	NA	This requirement is unnecessary in the DataConsensus application, as the data is submitted by the members and collected prior to the data requests.
17	Yes	This is fulfilled by the Duration constraint.
18	Yes	This is fulfilled by the `dpv:hasJustification` property in the duration constraint.
19	Yes	The third party confirms this when they submit the request.
20	Yes	This requirement is fulfilled by the purpose, selling data and selling insights constraints.
21	Yes	This is fulfilled by the technical organisation measures constraint.
22	Partial	The existence of a data protection impact assessment can be specified in the Technical Organisation Measures constraint but it could be argued that this should be its own constraint to ensure no ambiguity.
23	No	There are no properties or constraints to communicate this information in OCP. Implementing this requirement could be possible using the Risk Extension for DPV, however, given the depth of this extension, it was not attempted in this project.

Continuation of Table 5.2		
#	**Fulfilled?**	**Explanation**
24	Yes	This is fulfilled by both the Jurisdiction and the Third Country constraints in OCP.
25	Partial	There is no formal input or property that reflects this information. No suitable vocabulary within DPV was identified to represent this. However, the third parties are asked to include this information in their research description, as seen in Figure 4.10.
26	No	There are no suitable terms related to ethical approval in DPV. Because of this, the REC aspect of the data requests have not been implemented.
27	No	There are no suitable terms related to ethical approval in DPV. Because of this, the REC aspect of the data requests have not been implemented.
28	No	There are no suitable terms related to ethical approval in DPV. Because of this, the REC aspect of the data requests have not been implemented.
29	No	There are no suitable terms related to ethical approval in DPV. Because of this, the REC aspect of the data requests have not been implemented.
30	Yes	Instructions for withdrawing consent can be found in the footer of all pages for members.
31	No	Currently the system does not have a process to facilitate this requirement. As it requires re-identifying the member's data and clear lines of dialogue likely between the third parties and the admin, it has not been included in this initial proof of concept.
32	Partial	The third party confirms this when they submit the request.
33	Yes	Members can freely join the data cooperative and leave whenever they want.
34	Yes	As seen in Section 4.4.4, the constraints and information for all proposals are clearly communicated in a standardised way.
35	Yes	Since the members of the data cooperative are anonymised to the third parties, there can be no consequences against them.
36	Yes	As explained in Section 4.4.6, a member can easily leave the data cooperative, therefore withdrawing consent. The data collected is also stored pseudonymised, rather than anonymised, so the data can be easily removed from the pool.

Continuation of Table 5.2		
#	**Fulfilled?**	**Explanation**
37	Partial	All policies are recorded and kept, including agreements even if they have been revoked. It should be noted that agreements can be edited after the fact to remove members from the data subject list if they have left the cooperative. However, the copy of the consent is kept on the DataConsensus pod, though they are easily viewable by members on the user interface and the agreements.ttl in the pod is accessible by all members.
38	Yes	This is fulfilled as the decision process allows members to put forward other terms or recipients of the data.
End of Table 5.2		

Table 5.2: Evaluation of the DataConsensus application against the Data Policy Checklist, including an explanation for the fulfillment.

5.2 Community-Related Guidelines

This research extracted 14 guidelines related to building a fair and sustainable data commons from related literature and existing organisations, as well as a few requirements specified in the Data Act. These guidelines are broader than the criteria in the data policy checklist above. The literature covering data spaces, commons and cooperatives are particularly dispersed, as evident in the fact that the terminology is still unsettled. However, the literature does comprehensively address and share:

- Core elements and principles of data commons;
- Steps to follow for building data spaces;
- Lessons learnt from previous attempts to build data sharing initiatives.

By extracting these from the literature and synthesising them into a list of guidelines, they can be used as both guidance for developing tools for community data sharing applications, as well as an evaluation tool. The guidelines were extracted from 10 sources, including MyData Global [85], the Data Spaces Support Centre [43], the Open Future Foundation [100] and the International Data Spaces [72] . Additionally, requirements for data spaces specified in the Data Act [36] were combined with these guidelines, though these requirements are not yet in effect nor fully developed.

Notably, the principles extracted from MyData Global were geared towards individualistic data sharing, as opposed to community led data sharing [85]. Similarly, the guidance

from International Data Spaces and some other sources were focused on commercial data sharing, as opposed to data sharing in the public interest [72], [103]. As such, it was necessary to create a set of guidelines that was specifically targeting data commons with the relevant insights from the literature but excluding the unnecessary ones.

Additionally, while there is ample literature covering the broad principles of data commons [100], [43] and [53], there are fewer guidelines to transform these principles into action. Certainly, there are no guidelines that are as comprehensive as the literature mentioned above focused on other data sharing initiatives. This is a key reason behind this contribution, ensuring that there are comprehensive guidelines specifically for data commons, ensuring more coverage of all the operations of a data commons.

In terms of consensus regarding the guidelines, the vast majority of literature concur that trust and accountability is important, however, they are sometimes combined with transparency and traceability [100] [93]. The guidelines proposed below separated the two, since transparency and traceability not only help to build trust, but aids research and could even be legally required, which differentiates it from trust and accountability. Interestingly, the literature reflecting on previous data sharing initiatives and the lessons learnt [47] [42], advocate most strongly for data harmonisation, which was overlooked by the literature more focused on principles and governance. However, interoperability was also commonly mentioned throughout the literature, which makes sense given the political and regulatory environment.

One guideline overlooked by much of the literature was the final guideline regarding productive decision making. This is interesting since the literature that did include it was exploring semantic forms of communicating consent [61]. The gap in the literature may be explained by the other literature prioritising governance frameworks and interoperability over the tangible human experience - even if there is an incentive to participate if the decision-making process makes the system slow and frustrating, users will not engage. The decision making guideline is separate from the human centric approach, as you must balance empowering the individuals in the system while also ensuring the system remains productive for everyone else involved. This is a core goal of this project and is a key differentiator between these guidelines and existing guidelines focused on data spaces or individualistic data sharing.

1. **Build around a specific community:** The community is a core pillar of a data commons, as mentioned in Section 2.3.2 and multiple sources specify that when designing a data commons, it is necessary to clearly define the community it will be serving. This ensures that the commons is tailored to the values and interests of its potential data holders. [100] [47] [43]

2. **Identify the data's value proposition:** Similarly to the defined community, knowing the value of the data and who will be looking to access and use the data is crucial for designing a sustainable data commons. This also ensures that the data sharing serves a meaningful purpose. [103] [43]

3. **Ensure motivational incentives exist:** Incentives drive participation and engagement within the data commons. Whether financial or intangible, incentives should be aligned with the values of the community and encourage active contribution and collaboration. [43] [53]

4. **Establish governance and rules:** Outline the structure, roles and responsibilities of the stakeholders involved in the data commons. Additionally, the data commons should develop clear agreements that detail these rights and obligations. [42] [103] [43]

5. **Have a clearly defined mission and values:** A clear mission and set of values guide the direction and decision-making within the data commons. Having these clearly defined means fewer disagreements and disputes in the long term. [42]

6. **Promote shared understanding:** Simplifying the governance model as much as possible and creating a shared vocabulary should allow the stakeholders of the data commons to easily engage with commons. [42]

7. **Minimise the cost of participation:** In the same vein as the shared understanding, by minimising the cost (in time, money and resources) of participating in the data commons for all stakeholders, there should be more engagement and buy-in. This should also include reducing the barriers to accessing the data for data recipients. [47]

8. **Harmonise the data:** This is a time-consuming process and if done automatically, it can limit the type of data the commons stores. However, it is necessary to ensure the data is pseudonymised and it can make the data more valuable to data recipients. [47] [42] Furthermore, the Data Act requires data spaces to have a technical means to access the data, and detailed descriptions of the dataset content, use restrictions, licences, data collection methodology, data quality and uncertainty to the recipient. [36]

9. **Ensure interoperability:** Interoperability ensures seamless collaboration between different systems and tools within the data commons. It allows the system to integrate utilities and tools from other data commons, and enables the combination of different data commons with one another. [42] [85] [93]

10. **Build trust and accountability:** Trust is foundational to any community. Implementing transparent processes, clear governance, and accountability measures helps build confidence among stakeholders and encourages active participation. [42] [93] [61]

11. **Be transparent and traceable:** Following on from building trust and accountability, transparency, along with the ability to trace actions and changes, enhances accountability and builds trust within the community. [85] [61] [93] Furthermore, the Data Act will mandate data spaces to publicly share the data structures, formats and vocabularies they use. [36]

12. **Design with a human-centric approach:** Designing with the end-user in mind ensures that the data commons is accessible and user-friendly. This includes avoiding complex legal language. This promotes engagement and makes the commons more effective in serving its community. [85] [61] [93]

13. **Allow flexibility and scalability:** The system should be scalable and be able to dynamically onboard new participants. Additionally, the members should be able to negotiate and adjust the rules that govern the data requests. [72]

14. **Ensure productive decision making:** Efficient and timely decision-making processes prevent bottlenecks and obstructions, leading to more data sharing and a more engaged community. [61]

Table 5.3 describes how the DataConsensus application fulfils these guidelines. It should be noted that given the breadth of these guidelines, there are multiple ways to fulfil them and arguably always a way to improve how the system fulfils them.

Guideline	**Fulfillment**
1	The DataConsensus application was built with diabetes patients who use Nightscout in mind, as outlined in Section 3.1.
2	As outlined in the user stories, the blood sugar levels throughout the day can be used for multiple research objectives. Additionally, the other value of the data is the fact that it is standardised and will supposedly represent a large group of individuals.

Continuation of Table 5.3	
Guideline	**Fulfillment**
3	Responsibilities and rights of each actor within the DataConsensus are clearly defined as outlined in Section 3.2. Additionally, the OCP policies serve as the agreements that detail rights and obligations. However there are certain structures with the system that have not been clearly defined, this includes how admins are appointed or elected or how members can collectively revoke their consent.
4	As outlined in Section 2.2.2, researchers are incentivised to use the DataConsensus application since it streamlines the process of collecting data, which has been proven more difficult since HRR. As for the members, Section 3.1 highlights how people with chronic diseases are more likely to share their data for research, since it could benefit them and others with their disease.
5	Section 1.1 describes the motivations of this project and it generally served as the imagined mission and values for the DataConsensus application, however, this would need to be further refined with consultation from the defined community if the DataConsensus was ever going into production..
6	DataConsensus benefits from having developed its own profile and therefore there is a shared vocabulary for describing much of the system. While the decision process can seem complex, the BPMN diagram serves as a resource to promote understanding and any further simplification would likely negatively impact the democracy process or the ability for the members to collaborate and negotiate offers.
7	From a third party perspective, the cost of participation should be minimal as the key task will be filling out the policy builder. This builder was designed to not overwhelm the user by limiting the vocabulary to only the most relevant ones. As for members, the onboarding process is relevantly straightforward, similarly with the voting process.
8	The DataConsensus application pseudonymised and combines the data pool, though it should be noted that it does this by requiring a specific type and format of the data and this is a limitation of the system. Furthermore, the data can be easily accessed by any pod browser in accordance with the Data Act. However, aside from the use restrictions, the application doesn't inform the recipients of other key information on the data, including data collection methodology and data quality.

Continuation of Table 5.3	
Guideline	**Fulfillment**
9	By using ODRL and DPV, the system can be adapted to other linked data technologies, such as GraphDB and given the growth of other health related ontologies, there should be more opportunities to build interoperable features in the system.
10	The system is built on Solid, emphasising individual privacy and data control, which should foster trust. Privacy and security measures, detailed in Sections 5.3 and 5.4, further enhance this trust. However, the system lacks key mechanisms for accountability, particularly in holding admins responsible, or ensuring third-party compliance with agreements.
11	OCP, the other schemas and the code for the DataConsensus application are all publicly available. This means that anyone can review, critique and build on the existing system. Furthermore, the three policy types, along with the three status properties, makes the decision making process very straightforward.
12	The user design was an important consideration within the project. The OCP policies were transformed from their RDF format to simple easy-to-read formats, as seen in Section 4.4.
13	While the performance of the Solid application may decrease if there was significant growth in the application, it should remain operational. Additionally, DataConsensus easily onboards new members, adding them to the data pool, adding them as data subjects and notifying approved third parties of the new data. In terms of flexibility, the members can counter third parties with their own offers, though the application could provide even more customisation in regards to the constraints.
14	The application enforces time limits on the deliberation processes to ensure timely discussions and votes. Additionally, enabling comments allows members to easily communicate their thoughts and ideas with each other and the admin's ability to moderate means that these discussions stay civilised and spam-free.
End of Table 5.3	

Table 5.3: A table presenting how the DataConsensus fulfils the community related guidelines.

5.3 Security Considerations

All applications or servers available online have the risk of being attacked and Solid no different. Therefore, Solid is subject to the same security considerations found in HTTP/1.1. The Solid protocol will always encrypt data in transit from a pod to app and vice versa but another consideration is the pod. We will first discuss threats and migrations concerning app security and then discuss security concerns regarding pod providers. There are the standard threats against a web application, such as HTTP Request Header Injection Attacks, Man-in-the-middle (MITM) Attacks and Distributed Denial of Services (DDoS) Attacks and in an production environment, migrations to these risks need to be enacted. However there are additional security concerns when it comes to Solid applications which will be discussed below.

5.3.1 Origin Bypass

A new SOLID feature called "trusted apps" has a known weakness for origin validation. When a user grants access to their Pod to a Solid online app, the app Origin is added to the user's trusted apps list. The app is then given a token that allows it to interact with the Pod on the user's behalf, with just the Origin contained in those interactions being confirmed using the trusted applications list. Although browsers send the Origin with HTTPS requests, non-browser clients don't have to send an Origin. This means that if someone gets hold of the token meant for a trusted app, they could potentially use a non-browser client to bypass the Origin validation and gain unauthorised access to the user's Pod. Solutions are planned to improve the security of this vulnerability but it is currently standing [95].

5.3.2 Pod Providers and Security

Users will have a choice for their own Pod provider, and the level of security within the pod depends on the provider. Pod providers don't have to encrypt the data stored and it is up to the user to assess the pod provider's security measures before engaging them. While GDPR [39] requires appropriate security measures to protect personal data, including technical and organisational measures, the area of security compliance is still influx, and a significant amount of responsibility falls on the user. Inrupt Pod Spaces, inrupt.net, solidcommunity.net, solidweb.org, and trinpod.us, use.id, solidweb.me, and Data pod are the current Pod providers available to Solid developers and users. These pod providers have different security measures and not all pod providers are suitable for sensitive personal information. Both the solidcommunity.net and solidweb.org Pod

providers are a prototype implementation of the Solid server. Although they are fully functional servers, they lack stability and security. Similarly, solidweb.me is used for testing purposes [96].

As highlighted above, the choice of pod provider will have a significant impact on how secure the user data will be. Another option not discussed above is self-hosting, where a user can fully control the security and data privacy. But at this point, it is not a user-friendly choice and does require technical understanding.

5.4 Privacy Considerations

A significant amount of the privacy considerations were discussed in Section 2.2, while reviewing the current regulatory regime for data protection and privacy rights. However, there are some more specific privacy considerations in relation to this work's implementation.

5.4.1 Pseudo-Anonymous Data

The data shared to the third parties in this application will be pseudo-anonymous by attaching a hash identifier created from a member's data source. While this method reduces the risk of exposing sensitive personal information, the data is not considered anonymous and is still legally considered personal data. It is important to note that pseudo-anonymous data can still be re-identified if an actor got access to the user's data source URL and the hash secret. However, this is a trade-off so a user can withdraw their own data if they so wish.

5.4.2 Solid Privacy Concerns

Despite its ethos supporting individual data control, there are outstanding concerns surrounding privacy and Solid pods. In this section, the two most prominent concerns will be explained. The main takeaway that even with Solid pods, users must be vigilant and proactive when selecting their provider to best protect their data and privacy.

Migrating Data

A principle of the SOLID protocol is interoperability and since the data stored in Solid Pods is structured and according to Linked Data principles, Pod providers are completely neutral regarding the Pod content. This means that the same data structure should be supported by any Pod provider. However, if a user wants to leave a pod, the user's

ability to migrate their data will depend on the provider and their terms and conditions. Additionally, a user may not just want to leave a pod but also want to delete their pods and the data with it. The deletion procedure will depend on the pod provider, highlighting the importance of the pod provider selection [95].

Third parties Replicating Data

A key proposition of the Solid protocol is its rules for controlling access to the data. Users are able to grant and revoke permission to third-party apps to access data in his/her Pod and control exactly what information the third party has access to. However, if an authorised third-party had previously made a copy of the data and you later revoked that application's access to your data, the copy will not be deleted. Thus, users must be vigilant with the access they give to their pods [95].

5.5 Summary

This project overall passes its evaluation. It implements the user stories outlined in Section 3.3.1 and the use cases outlined too. It is fully compliant with nearly half of the legal requirements for the policies, however it is partially compliant with roughly another third of the requirements, where the information is present but it may be ambiguous. Most of the requirements that were not attempted were because the DPV vocabulary didn't have the specific coverage for those requirements. The community related guidelines were gathered from reviewed literature and these guidelines were generally followed with designing and building this application. Furthermore, this chapter reviews the privacy and security considerations for the application, outlining the risks and migrations. This chapter proposes two key contributions from this research, the data policy checklist and the community-related guidelines, which can be used by other data commons to evaluate their practices.

Chapter 6

Conclusions & Future Work

This project aimed to answer the question "How to enable data cooperatives to collaboratively deliberate and decide on interoperable data requests from third parties?". The driving force behind this project is the growing European effort to share data in a fair and privacy-preserving manner [34]. Data cooperatives provide an efficient way for members to pool and control their data. Building a tool that not only facilitates the democratic decision-making process but also enables members to engage with and adjust the data requests they receive should provide more control to those members while encouraging more data sharing.

This project answered its research question by designing and developing the DataConsensus application. The DataConsensus application is a tool built on Solid and designed to manage data requests from third parties wishing to access the pooled CGM data of a group of Nightscout users. The design of this application was guided by having a defined community, identifying the actors and responsibilities and creating realistic user stories and use cases. These design choices were influenced by the literature review conducted in Chapter 2, which investigated the intersection of legislation, data commons and linked data. While the application is designed for the specific community and mission in mind, this project's contributions can be adapted to other scenarios in the future. This is thanks to the interoperability focus of this project.

The project managed this interoperability focus by utilising the Solid specification and linked data. Linked data and public ontologies foster interoperability thanks to their openness. This project proposes the ODRL Profile for Collective Policies (OCP), using ODRL and DPV ontologies, and building off the work from [33] and [91]. The OCP enables the easy traceability of the decision-making process, from request to access. The OCP is a public ODRL profile and can be used by other data cooperative tools so long as they facilitate linked data.

The application achieves the collaboratively deliberation by allowing members to propose counter-offers in response to data requests, adjusting the constraints and rules that govern the request to something more acceptable to the members. Additionally, the application allows members to comment and enable with each other directly below policies to encourage discussions and open deliberations.

This collaborative decision making process is outlined in the BPMN diagram presented in Chapter 3. This process was developed based on user stories extrapolated from related literature and real-world scenarios. This process was designed in conjunction with OCP and it can be adapted to work with other technologies aside from Solid, aiding further development in this area.

Based on the literature review in Chapter 2, key requirements and best practices involved in operating a data cooperative were extracted. Two checklists were then formulated from these insights, one for the required information for data requests and one for designing a fair and sustainable data cooperative. These checklists served as strict evaluation measures in this project, with the application at least partially fulfilling 80% of the information requirements and all of the community related items. These checklists identified areas of ambiguity in the legal compliance and areas where the governance of the community could be improved. Furthermore, future researchers and developers can utilise these checklists to evaluate and improve their own data cooperatives tools and designs.

Overall, this project has made significant contributions to the research area surrounding data cooperatives. It developed the DataConsensus application as a proof of concept, utilising the proposed decision-making framework to enable collaborative deliberation among data cooperative members and third parties. Additionally, it has contributed to the ODRL community with its OCP, ensuring the DataConsensus application is transparent, interoperable and traceable.

6.1 Future Work

This section will outline the areas and questions that can be further explored. Additionally, there are a number of features described in this section that would improve the DataConsensus application. This section won't focus on future work that could be pursued to improve the solid specification, which is briefly discussed in Section 4.6.

Firstly, because OCP and the proposed decision process are interoperable, a similar data cooperative tool could be developed using another technology other than Solid. Likewise, a hybrid approach combining Solid and a relational database as the data layer could be implemented and this might improve the overall performance of the application.

Comparing these technology alternatives may be an interesting proposition for other researchers.

Another interesting comparison that could be explored is the effectiveness and engagement of the different data cooperative frameworks discussed in this report. While this report does review existing data cooperatives frameworks, given the limited public information, this review was not in-depth and the community would benefit from more in-depth documentation and public feedback.

This project also would encourage the introduction of other alternative frameworks for data cooperatives. Examples of alternative frameworks include allowing members to vote for each individual part of a policy as suggested by [91] or allowing for direct communication between members and third parties. Additionally improvements to the decision process proposed in this work could be achieved by improving how third parties can loop through the system. Now is the time to experiment with the frameworks and decision processes of data cooperatives as the norms have not been established.

For DataConsensus to go to production, it would be necessary to engage with the Nightscout community in Ireland. An evaluation of its user experience and its effectiveness as a data cooperative tool should be pursued in connection with the community, though organising this evaluation could be difficult and costly. However, there is a lack of research into the deliberation processes for data sharing initiatives so having this user testing well-documented would be very beneficial to the research community.

The DataConsensus application is also lacking some required and useful features. In terms of requirements, improving the policy builder to avoid ambiguity surrounding recipients, data protection impact assessments and other missing requirements as outlined in Section 5.1. This can be completed by expanding the number of constraints in each policy. Additionally, the policies can be augmented further with DPV expansions could also be beneficial especially in regards to formalising risks. Finally, incorporating other vocabularies that covers ethical considerations and approval is also necessary work for the DataConsensus application in the future. One vocabulary that could be considered is the Data Use Ontology (DUO) which provides vocabulary related to data discovery and responsible sharing of genomics data [81], but others may also exist. Other vocabularies that could be incorporated into OCP is the Common Impact Data Standard Ontology [21] or PROV-O [7].

Another functionality that should be developed is allowing third parties to submit joint requests that would cater to use cases where joint data controllers are involved. Other customization could be added to the system, allowing the data communities to decide what constraints to offer third parties or to allow different data pools to be selected. Additionally, there are a number of defaults within the system such as the deliberation times and

communities should be empowered to change these defaults. Other rules could be added to the system, such as changing the voting system for the multiple offers from majority voting to single transferable vote or the calculation of the threshold value. These additional features would greatly improve a community's ability to govern itself and change over time.

More accountability measures could be added to the system such as admin elections and repeal votes. Allowing users to elect admins or even depose admins could provide an additional layer of accountability in the system. By providing statistics and visualisations of votes, comments and approvals, further transparency and traceability could be incorporated into the system and aid accountability measures. Accountability features targeting third parties could involve regular communications updates and even post research reports. This may also create the opportunity for researchers to continue to work with the data community and expand on their work.

Some of the benefits of using publicly available ontologies are overlooked in this project, as the vocabulary terms provided to users in the policy builder are hard coded into the system. This is because of the filtering described in Section 3.5.1. However, the frontend was directly linked to the public ontologies, it would automatically update when these ontologies update, easily keeping up to date with changes in the data protection ecosystem.

A large area to explore related to the DataConsensus application is monetization and revenue sources. Monetary exchanges or incentives were excluded from the scope of these projects, however, if the DataConsensus framework was expanded to larger data communities, providing monetary incentives may be the most efficient way to recruit and retain members.

As the regulatory regimes of the Data Act and the Data Governance Act, along with any other relevant legislation, become more defined, reviewing the checklist and guidelines proposed in Chapter 5 will be necessary to stay on top of legal compliance and market trends. Given these emerging legal environments, this report concludes that there will be ample need for further work in the realm of data sharing and collective data governance.

Bibliography

[1] Access policies: Universal api. Available at https://docs.inrupt.com/developer-tools/javascript/client-libraries/tutorial/manage-access-policies-universal/.

[2] Authenticate (node.js: Single-user app). Available at https://docs.inrupt.com/developer-tools/javascript/client-libraries/tutorial/authenticate-nodejs-script/.

[3] Authenticate (node.js web server). Available at https://docs.inrupt.com/developer-tools/javascript/client-libraries/tutorial/authenticate-nodejs-web-server/.

[4] Inrupt podspaces. Available at https://start.inrupt.com/profile.

[5] Manage access to data (acp). Available at https://docs.inrupt.com/developer-tools/javascript/client-libraries/tutorial/manage-acp/.

[6] Manage access to data (wac). Available at https://docs.inrupt.com/developer-tools/javascript/client-libraries/tutorial/manage-wac/.

[7] (2013). Prov-o: The prov ontology. Available at https://www.w3.org/TR/2013/REC-prov-o-20130430/#prov-o-at-a-glance.

[8] (2016). Linked data. Available at https://www.w3.org/wiki/LinkedData.

[9] (2016). Nightscout ireland facebook group. Available at https://www.facebook.com/groups/624985814344432/.

[10] (2020). *The European data strategy: shaping Europe's digital future.* European Commission. Directorate General for Communication.

[11] Abraham, R., Schneider, J., and vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49:424–438.

[12] Agarwal, S., Steyskal, S., Antunovic, F., and Kirrane, S. (2018). Legislative compliance assessment: framework, model and gdpr instantiation. In *Privacy Technologies and Policy: 6th Annual Privacy Forum, APF 2018, Barcelona, Spain, June 13-14, 2018, Revised Selected Papers 6*, pages 131–149. Springer.

[13] Baloup, J., Bayamlıoğlu, E., Benmayor, A., Ducuing, C., Dutkiewicz, L., Lalova, T., Miadzvetskaya, Y., and Peeters, B. (2021). White paper on the data governance act. *SSRN Electronic Journal*.

[14] Becker, R., Thorogood, A., Bovenberg, J., Mitchell, C., and Hall, A. (2022). Applying gdpr roles and responsibilities to scientific data sharing. *International Data Privacy Law*, 12(3):207–219.

[15] Bietti, E., Etxeberria, A., Mannan, M., and Wong, J. (2021). Data cooperatives in europe: A legal and empirical investigation. *White Paper Created as Part of The New School's Platform Cooperativism Consortium and Harvard University's Berkman Klein Center for Internet & Society Research Sprint*.

[16] Board, H. R. (2021). Gdpr and health research.

[17] CFRI (2022a). Use of data. Available At `https://cfri.ie/use-of-data/`.

[18] CFRI (2022b). Welcome to the cf registry of ireland. Available At `https://cfri.ie/`.

[19] Clarke, N., Vale, G., Reeves, E. P., Kirwan, M., Smith, D., Farrell, M., Hurl, G., and McElvaney, N. G. (2019). GDPR: an impediment to research? *Irish Journal of Medical Science (1971 -)*, 188(4):1129–1135.

[20] Coburn, A., Pavlik, e., Zagidulin, D., Migus, A., and White, R. (2022). Solid-oidc. Available At `https://solidproject.org/TR/oidc#concepts-webids`.

[21] Common Approach (2023). Common impact data standard version 2.1.

[22] Council of the EU and the European Council (2022). Council approves data governance act. *Council of the EU and the European Council*.

[23] Council of the EU and the European Council (2023). Data act: Council and parliament strike a deal on fair access to and use of data. *Council of the EU and the European Council*.

[24] Courbier, S., Dimond, R., and Bros-Facer, V. (2019). Share and protect our health data: an evidence based approach to rare disease patients' perspectives on data sharing and data protection - quantitative survey and recommendations. *Orphanet Journal of Rare Diseases*, 14(1).

[25] Data Protection Commission (2023). Special category data. Available at `https://www.dataprotection.ie/en/organisations/know-your-obligations/lawful-processing/special-category-data#:~:text=Special%20categories%20of%20personal%20data&text=Personal%20data%20revealing%20racial%20or,Trade%20union%20membership`. [Accessed 09-08-2023].

[26] De Bot, D. and Haegemans, T. (2021). Data sharing patterns as a tool to tackle legal considerations about data reuse with solid: Theory and applications in europe. *Digita Research Reports*.

[27] De Vos, M., Kirrane, S., Padget, J., and Satoh, K. (2019). Odrl policy modelling and compliance checking. In Fodor, P., Montali, M., Calvanese, D., and Roman, D., editors, *Rules and Reasoning*, pages 36–51, Cham. Springer International Publishing.

[28] Department of Health (2018). Guidance on information principles for informed consent for the processing of personal data for health research. Available at `https://www.hrb.ie/fileadmin/1._Non-plugin_related_files/RSF_files/GDPR_guidance_for_researchers/Health_Research_Information_Principles.pdf`.

[29] Department of Health (2021). Guidance on explicit consent amendment to the health research regulations. Available at `https://www.gov.ie/pdf/?file=https://assets.gov.ie/120262/df5541bb-1682-4f90-a740-69cd9178fc39.pdf#page=null`.

[30] Esteves, B., Pandit, H. J., and Rodríguez-Doncel, V. (2021). Odrl profile for expressing consent through granular access control policies in solid. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 298–306. IEEE.

[31] Esteves, B., Pandit, H. J., and Rodríguez-Doncel, V. (2022a). Odrl profile for access control v0.2. Available At `https://w3id.org/oac`.

[32] Esteves, B. and Rodríguez-Doncel, V. (2022). Analysis of ontologies and policy languages to represent information flows in gdpr. *Semantic Web*, (Preprint):1–35.

[33] Esteves, B., Rodríguez-Doncel, V., Pandit, H. J., Mondada, N., and McBennett, P. (2022b). Using the odrl profile for access control for solid pod resource governance. In Groth, P., Rula, A., Schneider, J., Tiddi, I., Simperl, E., Alexopoulos, P., Hoekstra,

R., Alam, M., Dimou, A., and Tamper, M., editors, *The Semantic Web: ESWC 2022 Satellite Events*, pages 16–20, Cham. Springer International Publishing.

[34] European Commission (2020). A european strategy for data. Available At https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066.

[35] European Commission (2023). European data strategy. Available at https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en. [Accessed 09-08-2023].

[36] European Commission and Directorate-General for Communications Networks, Content and Technology (2022). Proposal for a regulation of the european parliament and of the council on harmonised rules on fair access to and use of data (data act). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A68%3AFIN.

[37] European Commission Staff (2022). Commission staff working document on common european data spaces.

[38] European Organisation for Rare Diseases (EURORDIS) (2023). Rare diseases amp; ehds: Finding a delicate balance between research needs and patient rights. Available At https://epha.org/rare-diseases-ehds-finding-a-delicate-balance-between-research-needs-and-patient#:~:text=According%20to%20EURORDIS%20survey%2C%2097,on%20diseases%20other%20than%20theirs.

[39] European Parliament and Council of the European Union (2016). Consolidated text: Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) (text with eea relevance). Available at https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04.

[40] European Parliament and Council of the European Union (2022). Regulation (eu) 2022/868 of the european parliament and of the council of 30 may 2022 on european data governance and amending regulation (eu) 2018/1724 (data governance act) (text with eea relevance). Available At https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868.

[41] Florea, M. and Esteves, B. (2023). I consent to these terms': A legal and technical approach for obtaining valid consent in solid.

[42] Gagnon-Turcotte, S., Sculthorp, M., and Coutts, S. (2021). Digital data partnerships: Building the foundations for collaborative data governance in the public interest.

[43] García Robles, A., Peeters, B., Otto, B., Stolwijk, C., Pezuela, C., Curry, E., Fernandes, E., Berkers, F., Korhonen, H., Hierro, J., and et al. (2023). Starter kit for data space designers v1. Available at https://dssc.eu/space/SK/29523973/Starter+Kit+for+Data+Space+Designers+%7C+Version+1.0+%7C+March+2023?attachment=%2Frest%2Fapi%2Fcontent%2F29523973%2Fchild%2Fattachment%2Fatt110592007%2Fdownload&type=application%2Fpdf&filename=DSSC-Starterkit-Version-1.0.pdf.

[44] Government of Ireland (2021). S.I. No. 18/2021 - Data Protection Act 2018 (Section 36(2)) (Health Research) (Amendment) Regulations 2021. Available at https://www.irishstatutebook.ie/eli/2021/si/18/made/en/print. [Accessed 09-08-2023].

[45] Government of Ireland (2022). Available At https://www.gov.ie/en/press-release/5f099-government-approves-the-drafting-of-ground-breaking-legislati#:~:text=A%20widely%20used%20definition%20is,with%20Co%2Doperatives%20in%20Ireland.

[46] Greshake Tzovaras, B., Angrist, M., Arvai, K., Dulaney, M., Estrada-Galiñanes, V., Gunderson, B., Head, T., Lewis, D., Nov, O., Shaer, O., Tzovara, A., Bobe, J., and Price Ball, M. (2019). Open humans: A platform for participant-centered research and personal data exploration. *GigaScience*, 8(6):giz076.

[47] Grossman, R. L. (2023). Ten lessons for data sharing with a data commons. *Scientific Data*, 10(1).

[48] Haeck, P. (2023). Europe's new data law, explained. *Politico*.

[49] Hardinges, J. (2018). How we're exploring the definition of a data trust. Available at https://theodi.org/article/how-were-exploring-the-definition-of-a-data-trust/.

[50] Ho, C.-H. and CHUANGT, T.-R. (2019). 12: Governance of communal data sharing. *Good data*, page 202.

[51] HUS Helsinki University Hospital and SITRA (2019). Aiming for easy and safe remote glucose monitoring for diabetic

childrennbsp;. Available at https://www.hus.fi/sv/aktuellt/aiming-easy-and-safe-remote-glucose-monitoring-diabetic-children.

[52] Iannella, R., Monegraph, r@iannel.la, Villata, S., INRIA, and serena.villata@inria.fr (2018). Odrl information model 2.2. Available at https://www.w3.org/TR/odrl-model/#constraint.

[53] Ideas for Change (2016). Saluscoop - towards citizen governance and management of health data. Available at https://static1.squarespace.com/static/5f232e29457c080a19e70bf5/t/5f8a1e40ce7cf718b1d0ebb0/1602887251669/saluscoop_informe.pdf.

[54] Insurance Ireland (2022). Insurance ireland guidance on data protection requirements for insurers when handling personal data (june 2022). Available at https://www.insuranceireland.eu/media/documents/FINAL%20Guidance%20on%20Data%20Protection%20Requirements%20for%20Insurers%20When%20Handling%20Personal%20Data%20June%202022.pdf.

[55] International Data Spaces (2023a). Available At https://internationaldataspaces.org/we/.

[56] International Data Spaces (2023b). 1.2 purpose and structure of the reference architecture. Available At https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/introduction/1_1_goals_of_the_international_data_spaces/1_2_purpose_and_structure_of_the_document.

[57] JacobMcConomy, Jeffz, and Cldwalker (2019). Using sparql on solid data. Available at https://forum.solidproject.org/t/using-sparql-on-solid-data/1335.

[58] Janssen, H. and Singh, J. (2022). Data intermediary. *Internet Policy Review*, 11(1).

[59] JornWildt and Kidehen (2018). Fun-fact - using sparql to query the type registry. Available at https://forum.solidproject.org/t/fun-fact-using-sparql-to-query-the-type-registry/776.

[60] Kariotis, T., Ball, M. P., Greshake Tzovaras, B., Dennis, S., Sahama, T., Johnston, C., Almond, H., and Borda, A. (2020). Emerging health data platforms: From individual control to collective data governance. *Data amp; Policy*, 2:e13.

[61] Kurteva, A., Chhetri, T. R., Pandit, H. J., and Fensel, A. (2021). Consent through the lens of semantics: State of the art survey and best practices. *Semantic Web*, (Preprint):1–27.

[62] Lewis, D. (2017). Ns-data-types.md. Available at https://github.com/danamlewis/OpenHumansDataTools/blob/master/NS-data-types.md.

[63] Linebaugh, K. (2014). Citizen hackers tinker with medical devices. Available at https://www.wsj.com/articles/citizen-hackers-concoct-upgrades-for-medical-devices-1411762843?tesla=y.

[64] McGlinn, K., Rutherford, M. A., Gisslander, K., Hederman, L., Little, M. A., and O'Sullivan, D. (2022). Fairvasc: A semantic web approach to rare disease registry integration. *Computers in Biology and Medicine*, 145:105313.

[65] Mee, B., Kirwan, M., Clarke, N., Tanaka, A., Manaloto, L., Halpin, E., Gibbons, U., Cullen, A., McGarrigle, S., Connolly, E. M., Bennett, K., Gaffney, E., Flanagan, C., Tier, L., Flavin, R., and McElvaney, N. G. (2020). What GDPR and the health research regulations (HRRs) mean for ireland: a research perspective. *Irish Journal of Medical Science (1971 -)*, 190(2):505–514.

[66] Micheli, M., Ponti, M., Craglia, M., and Suman, A. B. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, 7(2):2053951720948087.

[67] MIDATA (2019). Articles of association of midata genossenschaft. Available At https://www.midata.coop/wp-content/uploads/2019/08/MIDATA_Statuten_20190626_EN.pdf.

[68] Midata Coop (2020). Midata coop. Available At https://www.midata.coop/.

[69] MIDATAcooperative (2023). Midatacooperative/open-midata-server: Open midata server. Available At https://github.com/MIDATAcooperative/open-midata-server.

[70] Mosely, M., Brackett, M., and Earley, S. (2009). The dama guide to the data management body of knowledge (dama-dmbok guide).

[71] MyData (2022). Suggested amendments to the eu data act by mydata global. Available At https://www.mydata.org/2022/10/25/amendments-to-the-eu-data-act/.

[72] Nagel, L. and Lycklama, D. (2021). Design principles for data spaces.

[73] Neville, S. (2022). Are data trusts a suitable stewardship model for the developing world? *Financial Times*.

[74] nodeSolidServer and NoelDeMartin (2018). Support sparql get requests as documented on solid-spec · issue 962 · nodesolidserver/node-solid-server. Available at https://github.com/nodeSolidServer/node-solid-server/issues/962.

[75] Oelsner, N. (2023). Explainer: Everything you need to know about the european health data space. *Euronews*.

[76] Open Data Institute (2019). The data spectrum. Available At https://www.theodi.org/about-the-odi/the-data-spectrum/.

[77] Open Future (2021). European Parliament misses the opportunity to build the Data Commons – Open Future. Available At https://openfuture.eu/blog/european-parliament-misses-the-opportunity-to-build-the-data-commons-2/. [Accessed 16-08-2023].

[78] Open Future Foundation (2023). Open future foundation. Available At https://openfuture.eu/.

[79] Pandit, H. (2022). DPV v1 Release — Data Privacy Vocabularies and Controls Community Group. Available at https://www.w3.org/community/dpvcg/2022/12/05/dpv-v1-release/. [Accessed 15-08-2023].

[80] Pandit, H. J. (2023). Making sense of solid for data governance and gdpr. *Information*, 14(2):114.

[81] Pandit, H. J. and Esteves, B. (2023). Enhancing data use ontology (duo) for health-data sharing by extending it with odrl and dpv. *Preprint on webpage at https://www.semantic-web-journal. net/system/files/swj3127. pdf*.

[82] Pavel, V. (2021). Exploring legal mechanisms for data stewardship. *Ada Lovelace Institute and UK AI Council*.

[83] Pierce, D. (2023). Can activitypub save the internet? Available at https://www.theverge.com/2023/4/20/23689570/activitypub-protocol-standard-social-network.

[84] Pištorová, B. and Plevák, O. (2023). Stakeholders doubtful eu health data space will launch on schedule. *EURACTIV*.

[85] Poikola, A., Kaplan, D., Mällo, T., and MyData (2017). Mydata declaration. Available at https://www.mydata.org/participate/declaration/.

[86] Renyuneyun, ThisIsMissEm, Mrkvon, Bourgeoa, and Hochstenbach (2023). How to authenticate an app with its own identity (e.g. webid)? Available at https://forum.solidproject.org/t/how-to-authenticate-an-app-with-its-own-identity-e-g-webid/5946/5.

[87] Richardson, D. A. (2023). Nightscout loader home. Available at https://david50703.wixsite.com/nightscoutloader, journal=nightscoutloader.

[88] S-P, Jeffz, and Bourgeoa (2021). Sparql update - general questions. Available at https://forum.solidproject.org/t/sparql-update-general-questions/4935.

[89] SalusCoop (2020a). [eng] standard consent salus cg license. Available At https://www.saluscoop.org/licencia.

[90] SalusCoop (2020b). License. Available At https://www.saluscoop.org/licencia.

[91] Selvakumar, L. (2022). Master's thesis, Trinity College Dublin.

[92] Sensotrend (2023). Available At https://www.sensotrend.com/.

[93] SITRA (2021). Principles-based frameworks and tools for fair data economy. Available At https://www.sitra.fi/app/uploads/2021/10/2-pager-principles-based-frameworks-and-tools-nov-2021.pdf.

[94] Solano, J. L., de Souza, S., Martin, A., and Taylor, L. (2022). Governing data and artificial intelligence for all: models for sustainable and just data governance.

[95] Solid Project. Available at https://solidproject.org/faqs.

[96] Solid Project. Get a pod. Available At https://solidproject.org/users/get-a-pod.

[97] Spencer, K., Sanders, C., Whitley, E. A., Lund, D., Kaye, J., and Dixon, W. G. (2016). Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: A qualitative study. *Journal of Medical Internet Research*, 18(4):e66.

[98] Starzak, J. (2019). Available at https://jstarpl.github.io/nightscout-exporter/.

[99] Stolton, S. (2020). New eu data brokers won't have to be european, commission says. *EURACTIV*.

[100] Tarkowski, A. and Zygmuntowski, J. J. (2022). *DATA COMMONS PRIMER*. Open Future.

[101] Taubman-Bassirian, T. and Montezuma, L. A. (2020). Available at https://iapp.org/news/a/how-to-avoid-consent-fatigue/.

[102] TCD Data Protection (2021). Tcd data protection handbook. Available at https://www.tcd.ie/dataprotection/assets/docs/dataprotectionhandbook/DP_Handbook_09032021.pdf.

[103] The GovLab (2023). Designing a data collaborative. Available At https://datacollaboratives.org/canvas.html.

[104] The Nightscout Data Commons Committee and The Nightscout Foundation (2017). Nightscout data commons. Available at https://www.openhumans.org/activity/nightscout-data-commons/.

[105] The Nightscout Foundation (2018). The nightscout foundation. Available at https://www.nightscoutfoundation.org/.

[106] The Nightscout Project (2023). Available at http://www.nightscout.info/.

[107] Turkmayali, A. (2023). Navigating personal data sharing: Opportunities and concerns. Available At https://internationaldataspaces.org/navigating-personal-data-sharing-opportunities-and-concerns/.

[108] van Geuns, J. and Brandusescu, A. (2020). What does it mean?: Shifting power through data governance. Available at https://foundation.mozilla.org/en/data-futures-lab/data-for-empowerment/shifting-power-through-data-governance/#what-is-a-data-collaborative.

[109] Veil, W. (2022). Data altruism: How the eu is screwing up a good idea. Technical report.

[110] Verhulst, S. G., Young, A., Winowatan, M., and Zahuranec, A. J. (2019). Leveraging private data for public good. *Govlab*.

[111] VHI (2023). Graduate testimonials. Available at https://www1.vhi.ie/about/careers/graduate-programme/testimonials.

[112] Vogelezang, F. (2022). A closer look at data intermediaries and the risk of platformization. Available at https://openfuture.eu/blog/a-closer-look-at-data-intermediaries-and-the-risk-of-platformization/.

[113] Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, EASE '14, New York, NY, USA. Association for Computing Machinery.

Appendix A

Use Of AI

The AI tools used for this project were GitHub Copilot's student subscription and the ChatGPT's premium subscription. The GitHub Copilot tool was active throughout the coding process but only its autocompleting features were utilised, not the comment prompting feature. The use of the GitHub Copilot was mostly motivated by curiousity and a desire to use the tools that other software engineers are using. Once installed, the extension quickly faded into the background, operating like any other extensions that helps with formatting or debugging.

ChatGPT was used in the report-writing, in particular during parts of Chapter 2 and Chapter 3. The tool was used to help with turns of phrase and expressions. For example, the first draft of the selling data constraint description in Chapter 3 was long and vague, specifically this sentence - "The negative definition is as beneficial in making the policy clear as the positive." ChatGPT suggested revised phrasing that contained the word "ambiguity". This was overlooked in the first draft but was suitable in a number of other parts of the report and added to the clarity and conciseness of the report. This greatly helped the writing process. Interestingly, without ChatGPT, more help with phrasing would have been requested from my family. Therefore, ChatGPT not only saved me the time and distraction it would take to reach other to my family, but also the hassle for my family.

Appendix B

ODRL Profile for Collective Policies (OCP) in Turtle Format

This appendix presents an example of each type of policy in the OCP in turtle format. The prefixes and namespaces used throughout the project are listed in Table B.1.

Figure B.1 is the main instance of the request policy belonging to the Insurance Company user story in Section 3.3.1. Figure B.2 is the permission instance of the OCP request. Figure B.3 includes the code for the first four constraints, including Purpose, Organisation, Selling Data and Selling Insights constraints. Figure B.4 covers the final four constraints, including the Duration, Technical Organisational Measures, Recipients and Jurisdiction constraints.

Figure B.5 is the main instance of the offer policy belonging to the Non-Profit Organisation user story in Section 3.3.2. This is followed by Figures B.6, B.7, B.8, which reflect the other instances linked to the main policy instance.

Finally, an example of the agreement policy can be seen in Figures B.5, B.6, B.7 and B.8. This example belongs to the Academic Researcher user story in Section 3.3.1.

Prefix	**Namespace**
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
owl	http://www.w3.org/2002/07/owl#
xsd	http://www.w3.org/2001/XMLSchema#
dct	http://purl.org/dc/terms/
foaf	http://xmlns.com/foaf/0.1/
odrl	http://www.w3.org/ns/odrl/2/
dpv	https://w3id.org/dpv#
dpv-legal	https://www.w3id.org/dpv/dpv-legal#
oac	https://w3id.org/oac#
ocp	https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schemas/ocp.ttl#
user	https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schemas/user#
project	https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schemas/project#
vote	https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schema/vote#
comment	https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd036/app/schema/comment#

Table B.1: All prefixes and namespaces used in DataConsensus

```
request:72013827-a31e-4a27-a7cb-0b86fa9ddf15
        rdf:type                    odrl:Request ;
        dct:
creator                 <https://id.inrupt.com/InsuranceCompany> ;
        dct:isPartOf                project:82e7f5d0-662b-4ec0-adb1-b3e52ff
32a03 ;
        dct:issued                  "2023-06-09T14:05:13"^^xsd:dateTime ;
        odrl:permission             request:72013827-a31e-4a27-a7cb-0b86fa9
ddf15_permission ;
        odrl:profile                oac:, ocp: ;
        odrl:uid                    request:72013827-a31e-4a27-a7cb-0b86fa9
ddf15 ;
        ocp:hasJustification        "This is a justification" ;
        ocp:hasConsequence          "This is a consequence" ;
        ocp:adminApproved       ocp:Blocked ;
        ocp:memberApproved      ocp:Rejected ;
        ocp:thirdPartyApproved  ocp:Approved .

```

Figure B.1: The main instance of the OCP request policy belonging to the Insurance Company user story

```
14
15 request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_permission
16         rdf:type         odrl:Permission ;
17         odrl:action      dpv:Use , dpv:Remove , dpv:Transform , dpv:
   Store , dpv:Copy ;
18         odrl:assignee    <https://id.inrupt.com/InsuranceCompany> ;
19         odrl:assigner    <https://id.inrupt.com/DataConsensus> ;
20         odrl:constraint
21             request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_purposeConstraint,
22             request:72013827-a31e-4a27-a7cb-0b86fa9ddf15
   _organisationConstraint,
23             request:72013827-a31e-4a27-a7cb-0b86fa9ddf15
   _sellingDataConstraint,
24             request:72013827-a31e-4a27-a7cb-0b86fa9ddf15
   _sellingInsightsConstraint,
25             request:72013827-a31e-4a27-a7cb-0b86fa9ddf15
   _durationConstraint,
26             request:72013827-a31e-4a27-a7cb-0b86fa9ddf15
   _techOrgMeasureConstraint,
27             request:72013827-a31e-4a27-a7cb-0b86fa9ddf15
   _recipientConstraint,
28             request:72013827-a31e-4a27-a7cb-0b86fa9ddf15
   _jurisdictionConstraint;
29         odrl:
   target      <https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868cd03
   6/app/data/Resource.csv> .
30
```

Figure B.2: The permission instance of the OCP request policy belonging to the Insurance Company user story

```
request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_purposeConstraint
        rdf:type            odrl:Constraint ;
        odrl:leftOperand    oac:Purpose ;
        odrl:operator       odrl:isA ;
        odrl:rightOperand   dpv:CommercialResearch .

request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_organisationConstraint
        rdf:type            odrl:Constraint ;
        odrl:leftOperand    ocp:Organisation ;
        odrl:operator       odrl:isA ;
        odrl:rightOperand   dpv:ForProfitOrganisation .

request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_sellingDataConstraint
        rdf:type            odrl:Constraint ;
        odrl:leftOperand    oac:Purpose ;
        odrl:operator       oac:isNotA ;
        odrl:rightOperand   dpv:SellDataToThirdParties .

request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_sellingInsightsConstraint
        rdf:type            odrl:Constraint ;
        odrl:leftOperand    oac:Purpose ;
        odrl:operator       odrl:isA ;
        odrl:rightOperand   dpv:SellInsightsToThirdParties .
```

Figure B.3: The Purpose, Organisation, SellingData and SellingInsights constraints of the OCP request policy belonging to the Insurance Company user story

```
54
55 request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_durationConstraint
56         rdf:type           odrl:Constraint ;
57         ocp:hasJustification       "This is a justification" ;
58         odrl:leftOperand   ocp:UntilTimeDuration ;
59         odrl:operator      odrl:eq ;
60         odrl:rightOperand  "2024-12-31T00:00:00"^^xsd:dateTime .
61
62 request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_techOrgMeasureConstraint
63         rdf:type           odrl:Constraint ;
64         odrl:leftOperand   oac:TechnicalOrganisationalMeasure ;
65         odrl:operator      ocp:isAllOf ;
66         odrl:rightOperand  dpv:DataProtectionTraining , dpv:
   ConsultationWithDPO , dpv:IncidentManagementProcedures , dpv:
   CryptographicMethods , dpv:NeworkSecurityProtocols , dpv:
   CertificationSeal , dpv:EncryptionInTransfer , dpv:
   MultiFactorAuthentication , dpv:NDA , dpv:LoggingPolicies , dpv:
   CodeOfConduct , dpv:CybersecurityTraining , dpv:
   IncidentReportingCommunication , dpv:ComplianceMonitoring , dpv:
   EncryptionAtRest , dpv:PhysicalAssessControlMethod .
67
68 request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_recipientConstraint
69         rdf:type           odrl:Constraint ;
70         ocp:hasJustification       "This is a justification" ;
71         odrl:leftOperand   oac:Recipient ;
72         odrl:operator      ocp:isAnyOf ;
73         odrl:rightOperand  <https://id.inrupt.com/InsuranceCompany> .
74
75 request:72013827-a31e-4a27-a7cb-0b86fa9ddf15_jurisdictionConstraint
76         rdf:type           odrl:Constraint ;
77         odrl:leftOperand   ocp:hasJurisdiction ;
78         odrl:operator      oac:eq ;
79         odrl:rightOperand  dpv-legal:IE .
80
```

Figure B.4: The Duration, Technical Organisational Measures, Recipients and Jurisdiction constraints of the OCP request policy belonging to the Insurance Company user story

```
offer:ff284cf0-2c7b-4718-b261-698119c2408c
        rdf:type                    odrl:Offer ;
        dct:creator                 <https://id.inrupt.com/jeffwinger> ;
        dct:isPartOf                project:4ddb1306-df85-4730-8545-14973
be50cd2 ;
        dct:issued                  "2023-06-10T17:26:35"^^xsd:dateTime ;
        odrl:permission             offer:ff284cf0-2c7b-4718-b261-698119c
2408c_permission ;
        odrl:profile                oac:, ocp: ;
        odrl:uid                    offer:ff284cf0-2c7b-4718-b261-698119c
2408c ;
        ocp:hasJustification        "This is a justification" ;
        ocp:hasConsequence          "This is a consequence" ;
        ocp:adminApproved       ocp:Pending ;
        ocp:memberApproved      ocp:Approved ;
        ocp:thirdPartyApproved  ocp:Pending .

```

Figure B.5: The offer policy belonging to the Non-Profit Organisation user story

```

offer:ff284cf0-2c7b-4718-b261-698119c2408c_permission
        rdf:type         odrl:Permission ;
        odrl:action      dpv:Remove , dpv:Copy , dpv:Store , dpv:
Transform , dpv:Use ;
        odrl:assignee    <https://id.inrupt.com/thirdparty> ;
        odrl:assigner    <https://id.inrupt.com/DataConsensus> ;
        odrl:constraint  offer:ff284cf0-2c7b-4718-b261-698119c2408
c_purposeConstraint , offer:ff284cf0-2c7b-4718-b261-698119c2408
c_organisationConstraint , offer:ff284cf0-2c7b-4718-b261-698119c2408
c_sellingDataConstraint , offer:ff284cf0-2c7b-4718-b261-698119c2408
c_sellingInsightsConstraint,offer:ff284cf0-2c7b-4718-b261-698119c2408
c_durationConstraint , offer:ff284cf0-2c7b-4718-b261-698119c2408
c_techOrgMeasureConstraint , offer:ff284cf0-2c7b-4718-b261-698119c2408
c_recipientConstraint ,
        offer:ff284cf0-2c7b-4718-b261-698119c2408
c_jurisdictionConstraint ;
        odrl:
target      <https://storage.inrupt.com/b41a41bc-203e-4b52-9b91-4278868c
d036/app/data/Resource.csv> .

```

Figure B.6: The permission instance of the OCP offer policy belonging to the Non-Profit Organisation user story

```
23
24 offer:ff284cf0-2c7b-4718-b261-698119c2408c_purposeConstraint
25         rdf:type           odrl:Constraint ;
26         odrl:leftOperand   oac:Purpose ;
27         odrl:operator      odrl:isA ;
28         odrl:rightOperand  dpv:NonCommercialResearch .
29
30 offer:ff284cf0-2c7b-4718-b261-698119c2408c_organisationConstraint
31         rdf:type           odrl:Constraint ;
32         odrl:leftOperand   ocp:Organisation ;
33         odrl:operator      odrl:isA ;
34         odrl:rightOperand  dpv:NonProfitOrganisation .
35
36 offer:ff284cf0-2c7b-4718-b261-698119c2408c_sellingDataConstraint
37         rdf:type           odrl:Constraint ;
38         odrl:leftOperand   oac:Purpose ;
39         odrl:operator      oac:isNotA ;
40         odrl:rightOperand  dpv:SellDataToThirdParties .
41
42 offer:ff284cf0-2c7b-4718-b261-698119c2408c_sellingInsightsConstraint
43         rdf:type           odrl:Constraint ;
44         odrl:leftOperand   oac:Purpose ;
45         odrl:operator      oac:isNotA ;
46         odrl:rightOperand  dpv:SellInsightsFromData .
47
```

Figure B.7: The Purpose, Organisation, SellingData and SellingInsights constraints of the OCP offer policy belonging to the Non-Profit Organisation user story

```
47
48 offer:ff284cf0-2c7b-4718-b261-698119c2408c_durationConstraint
49         rdf:type           odrl:Constraint ;
50         ocp:hasJustification       "This is a justification" ;
51         odrl:leftOperand   ocp:UntilTimeDuration ;
52         odrl:operator      odrl:eq ;
53         odrl:rightOperand  "2024-11-08T00:00:00"^^xsd:dateTime .
54
55 offer:ff284cf0-2c7b-4718-b261-698119c2408c_techOrgMeasureConstraint
56         rdf:type           odrl:Constraint ;
57         odrl:leftOperand   oac:TechnicalOrganisationalMeasure ;
58         odrl:operator      ocp:isAllOf ;
59         odrl:rightOperand  dpv:ConsultationWithDPO , dpv:
   OperatingSystemSecurity , dpv:PasswordAuthentication , dpv:
   CodeOfConduct , dpv:IncidentReportingCommunication , dpv:
   EncryptionAtRest , dpv:EncryptionInTransfer , dpv:
   DataProtectionTraining .
60
61 offer:ff284cf0-2c7b-4718-b261-698119c2408c_recipientConstraint
62         rdf:type           odrl:Constraint ;
63         ocp:hasJustification       "This is a justification" ;
64         odrl:leftOperand   oac:Recipient ;
65         odrl:operator      ocp:isAnyOf ;
66         odrl:rightOperand
67             <https://id.inrupt.com/thirdparty> .
68
69 offer:ff284cf0-2c7b-4718-b261-698119c2408c_jurisdictionConstraint
70         rdf:type           odrl:Constraint ;
71         odrl:leftOperand   ocp:ocp:hasJurisdiction ;
72         odrl:operator      oac:eq ;
73         odrl:rightOperand  dpv-legal:IE .
```

Figure B.8: The Duration, Technical Organisational Measures, Recipients and Jurisdiction constraints of the OCP offer policy belonging to the Non-Profit Organisation user story

```
agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9
        rdf:type            odrl:Agreement ;
        dct:creator         <https://id.inrupt.com/dataconsensusadmin> ;
        dct:isPartOf        project:af993b9f-b8ba-4f3f-9f5e-26b6d37747
ae ;
        dct:issued          "2023-06-13T18:13:37"^^xsd:dateTime ;
        dct:references      request:12d6eb50-f6e2-488c-8923-fb0c4d47749
a ;
        odrl:permission     agreement:639727b3-c8ca-47ea-ba3e-51fc68d05
bd9_permission ;
        odrl:profile        oac: ;
        odrl:uid            agreement:639727b3-c8ca-47ea-ba3e-51fc68d05
bd9 ;
        ocp:adminApproved       ocp:Approved ;
        ocp:memberApproved      ocp:Approved ;
        ocp:thirdPartyApproved  ocp:Approved ;
        dpv:
hasDataSubject          <https://id.inrupt.com/DataConsensus> ;
        dpv:
hasJointDataController  <https://id.inrupt.com/DataConsensus> , <http
s://id.inrupt.com/DrJohnSmith> ;
        dpv:hasLegalBasis           dpv:Consent .

```

Figure B.9: The agreement policy belonging to the Academic Researcher user story

```
16
17 agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9_permission
18         rdf:type         odrl:Permission ;
19         odrl:action      dpv:Transform , dpv:Store , dpv:Use , dpv:
   Remove , dpv:Copy ;
20         odrl:assignee    <https://id.inrupt.com/DrJohnSmith> ;
21         odrl:assigner    <https://id.inrupt.com/DataConsensus> ;
22         odrl:constraint  agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9
   _purposeConstraint , agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9
   _organisationConstraint , agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9
   _sellingDataConstraint ,  agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9
   _sellingInsightsConstraint ,agreement:639727b3-c8ca-47ea-ba3e-51fc68d05
   bd9_durationConstraint ,agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9
   _techOrgMeasureConstraint , agreement:639727b3-c8ca-47ea-ba3e-51fc68d05
   bd9_recipientConstraint ,
23         agreement:ff284cf0-2c7b-4718-b261-698119c2408
   c_jurisdictionConstraint ;
24         odrl:target      dpvpd:MedicalHealth .
25
```

Figure B.10: The permission instance of the OCP agreement policy belonging to the Academic Researcher user story

```
25
26 agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9_purposeConstraint
27         rdf:type           odrl:Constraint ;
28         odrl:leftOperand   oac:Purpose ;
29         odrl:operator      odrl:isA ;
30         odrl:rightOperand  dpv:AcademicResearch .
31
32 agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9_sellingDataConstraint
33         rdf:type           odrl:Constraint ;
34         odrl:leftOperand   oac:Purpose ;
35         odrl:operator      odrl:isNotA ;
36         odrl:rightOperand  dpv:SellDataToThirdParties .
37
38 agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9_sellingInsightsConstraint
39         rdf:type           odrl:Constraint ;
40         odrl:leftOperand   oac:Purpose ;
41         odrl:operator      odrl:isNotA ;
42         odrl:rightOperand  dpv:SellInsightsFromData .
43
44 agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9_organisationConstraint
45         rdf:type           odrl:Constraint ;
46         odrl:leftOperand   oac:Organisation ;
47         odrl:operator      odrl:isA ;
48         odrl:rightOperand  dpv:AcademicScientificOrganisation .
49
```

Figure B.11: The Purpose, Organisation, SellingData and SellingInsights constraints of the OCP agreement policy belonging to the Academic Researcher user story

```
49
50 agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9_techOrgMeasureConstraint
51         rdf:type           odrl:Constraint ;
52         odrl:leftOperand   oac:TechnicalOrganisationalMeasure ;
53         odrl:operator      odrl:isAllOf ;
54         odrl:rightOperand  dpv:IncidentReportingCommunication , dpv:
   ComplianceMonitoring , dpv:EncryptionInTransfer , dpv:
   PasswordAuthentication , dpv:EncryptionAtRest , dpv:
   DataProtectionTraining , dpv:CodeOfConduct , dpv:ConsultationWithDPO .
55
56 agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9_recipientConstraint
57         rdf:type           odrl:Constraint ;
58         odrl:leftOperand   oac:Recipient ;
59         odrl:operator      odrl:isAnyOf ;
60         odrl:
   rightOperand  <https://id.inrupt.com/DrJohnSmith> , <https://id.inrupt.c
   om/DataConsensus> ,
   dpv:AcademicScientificOrganisation .
61
62 agreement:639727b3-c8ca-47ea-ba3e-51fc68d05bd9_durationConstraint
63         rdf:type           odrl:Constraint ;
64         odrl:leftOperand   dpv:UntilTimeDuration ;
65         odrl:operator      odrl:eq ;
66         odrl:rightOperand  "2023-12-31T00:00:00"^^xsd:dateTime .
67
68 agreement:ff284cf0-2c7b-4718-b261-698119c2408c_jurisdictionConstraint
69         rdf:type           odrl:Constraint ;
70         odrl:leftOperand   ocp:ocp:hasJurisdiction ;
71         odrl:operator      oac:eq ;
72         odrl:rightOperand  dpv-legal:IE .
```

Figure B.12: The Duration, Technical Organisational Measures, Recipients and Jurisdiction constraints of the OCP agreement policy belonging to the Academic Researcher user story

Appendix C

Supplemental Material

This appendix covers relevant material for anyone further developing the DataConsensus application.

C.1 Nightscout Data Example

	A	B	C
1	2017-07-06T19:56:43.365-0400	77	
2	2017-07-06T19:51:42.652-0400	80	
3	2017-07-06T19:46:44.154-0400	87	
4	2017-07-06T19:41:42.945-0400	99	
5	2017-07-06T19:36:42.385-0400	113	
6	2017-07-06T19:26:43.538-0400	119	
7	2017-07-06T19:21:42.708-0400	153	
8	2017-07-06T19:16:42.615-0400	164	
9	2017-07-06T19:11:43.374-0400	175	
10	2017-07-06T19:06:42.645-0400	186	
11	2017-07-06T19:01:43.356-0400	194	
12	2017-07-06T18:56:42.861-0400	197	
13	2017-07-06T18:51:42.880-0400	191	
14	2017-07-06T18:46:42.793-0400	176	
15	2017-07-06T18:41:43.072-0400	154	
16	2017-07-06T18:36:43.102-0400	128	
17	2017-07-06T18:31:42.968-0400	104	
18	2017-07-06T18:26:42.160-0400	89	
19	2017-07-06T18:21:43.189-0400	83	
20	2017-07-06T18:16:42.677-0400	77	
21	2017-07-06T18:11:42.747-0400	80	

Figure C.1: An example of the Nightscout's entries.csv from [62]

C.2 WebIDs and Pods created for this Project

WebID	Description
https://id.inrupt.com/dataconsensus	This is the main account linked to the solid pod that holds the application's data.
https://id.inrupt.com/dataconsensusadmin	This is the account for the admin and is solely used for authentication purposes. In the current iteration of the application, it is not possible to register a new admin user, they must be manually added in the admins.ttl file.
https://id.inrupt.com/thirdparty	This is the account for the test third party.
https://id.inrupt.com/odonneb4	This is the account used as the test member. It should be noted that within the pod linked to this account, there is the sampleData.csv file that serves as the data source. A copy of this sampleData.csv file can be found in the GitHub repo.

Table C.1: A table of the WebIDs created for the development of this project, along with descriptions for each WebID

C.3 Instructions for Setting Up the DataConsensus Application

To set up this application in a local environment, clone the GitHub repository into a directory of your choice. The base directory contains three folders:

1. DataConsensus_Backend;
2. DataConsensus_Frontend;
3. POD.

The latter contains the demo files for the DataConsensus application.

To set up your own version of the application, you will need to create a WebID and pod to serve as the DataConsensus pod. You can do that here: `https://start.inrupt.com/` Additionally, you will need to register an app to your new WebID and Pod. Take

note of the client id and secret. You can do that here: `https://login.inrupt.com/registration.html` Once you have a pod set up to serve as your DataConsensus pod, upload the files in the POD folder. You may want to change the admins.ttl file to include your personal WebID so you can access all aspects of the frontend. Furthermore, you will need to create an app password for an email of your choice. To create one using google, follow these instructions: `https://support.google.com/mail/answer/185833?hl=en`

The public GitHub repository is missing the .env file located in the DataConsensus_Backend directory. It contains a number of essential constants used through the application's backend. These include the URLs of the DataConsensus pod used in development and testing and other constants that cannot be shared. Therefore, you will need to create a new .env with new constants. A list of the constants to include can be seen in Table C.2.

In the directory where the cloned repo is located, open two terminals and in the first terminal, run the following commands:

```
cd DataConsensus_Backend
npm install
npm start
```

In the second terminal, run the following commands:

```
cd DataConsensus_Frontend
ng build
ng serve
```

C.4 Development Guide

C.4.1 Backend Development

This section will serve as a simple explanation of the overall structure and flow of the DataConsensus backend. The index.js file serves as the entry point to the Node.js application. It contains the code for the app authentication as discussed in Section 4.2 and uses the @inrupt/solid-client-authnnode library to manage the session. It also implements the Express.js web server with Cross-Origin Resource Sharing enabled and specifies a number of different routes for handling requests. The RESTful endpoints are divided across the following files located in the Routes directory: AuthRoute.js, CommentRoute.js, PolicyRoute.js, ProjectRoute.js, UserRoute.js, VoteRoute.js. Figure C.2 presents the API endpoints in the previously stated files.

Name	Example
FRONTEND	http://localhost:4200
HOSTNAME	http://localhost:
PORT	3000
API_URI	/api/v1
ODRL	http://www.w3.org/ns/odrl/2/0
DPV	"https://w3id.org/dpv#"
OAC	"https://w3id.org/oac#"
DPVLEGAL	"https://www.w3id.org/dpv/dpv-legal#"
USER	YOUR-POD-LOCATION/app/schemas/user.ttl
COMMENT	YOUR-POD-LOCATION/app/schemas/comment.ttl
OCP	YOUR-POD-LOCATION/app/schemas/ocp.ttl
PROJECT	YOUR-POD-LOCATION/app/schemas/project.ttl
VOTE	YOUR-POD-LOCATION/app/schemas/vote.ttl
MEMBER_LIST	YOUR-POD-LOCATION/app/members.ttl
THIRDPARTY_LIST	YOUR-POD-LOCATION/app/third-parties.ttl
ADMIN_LIST	YOUR-POD-LOCATION/app/admins.ttl
REQUESTS	YOUR-POD-LOCATION/app/requests.ttl
OFFERS	YOUR-POD-LOCATION/app/offers.ttl
AGREEMENTS	YOUR-POD-LOCATION/app/agreements.ttl
COMMENTS	YOUR-POD-LOCATION/app/comments.ttl
VOTES	YOUR-POD-LOCATION/app/votes.ttl
PROJECTS	YOUR-POD-LOCATION/app/projects.ttl
SECRET	A sha 256 secret for the hashing function
APP_CLIENT_ID	This is the client id you received when you registered your app
APP_CLIENT_SECRET	This is the client secret you received when you registered your app
EMAIL	An email of your choice
PASSWORD	The app password for the email of your choice
APP_OIDC_ISSUER	https://login.inrupt.com
RESOURCE_URL	YOUR-POD-LOCATION/app/pool/datapool.csv

Table C.2: Constants in the .env file used in the DataConsenus_Backend server

The classes used in the backend are defined within files in the Models directory. Figure C.3 presents the classes involved in this project. The CRUDService directory contains the five Service files that interact with the pod's data. Within the base directory, the HelperFunction.js file contains functions used repeatedly throughout the code, while the AccessControl.js file contains the grantAccess and removeAccess functions which use the universalAccess module in the @inrupt/solid-client library. Finally, there is a Logic directory, which contains the functionality for mailing notification and transforming the user data into pseudonymised data suitable for the data pool. The backend leverages environment variables for configuration, ensuring that the URLs and sensitive information used

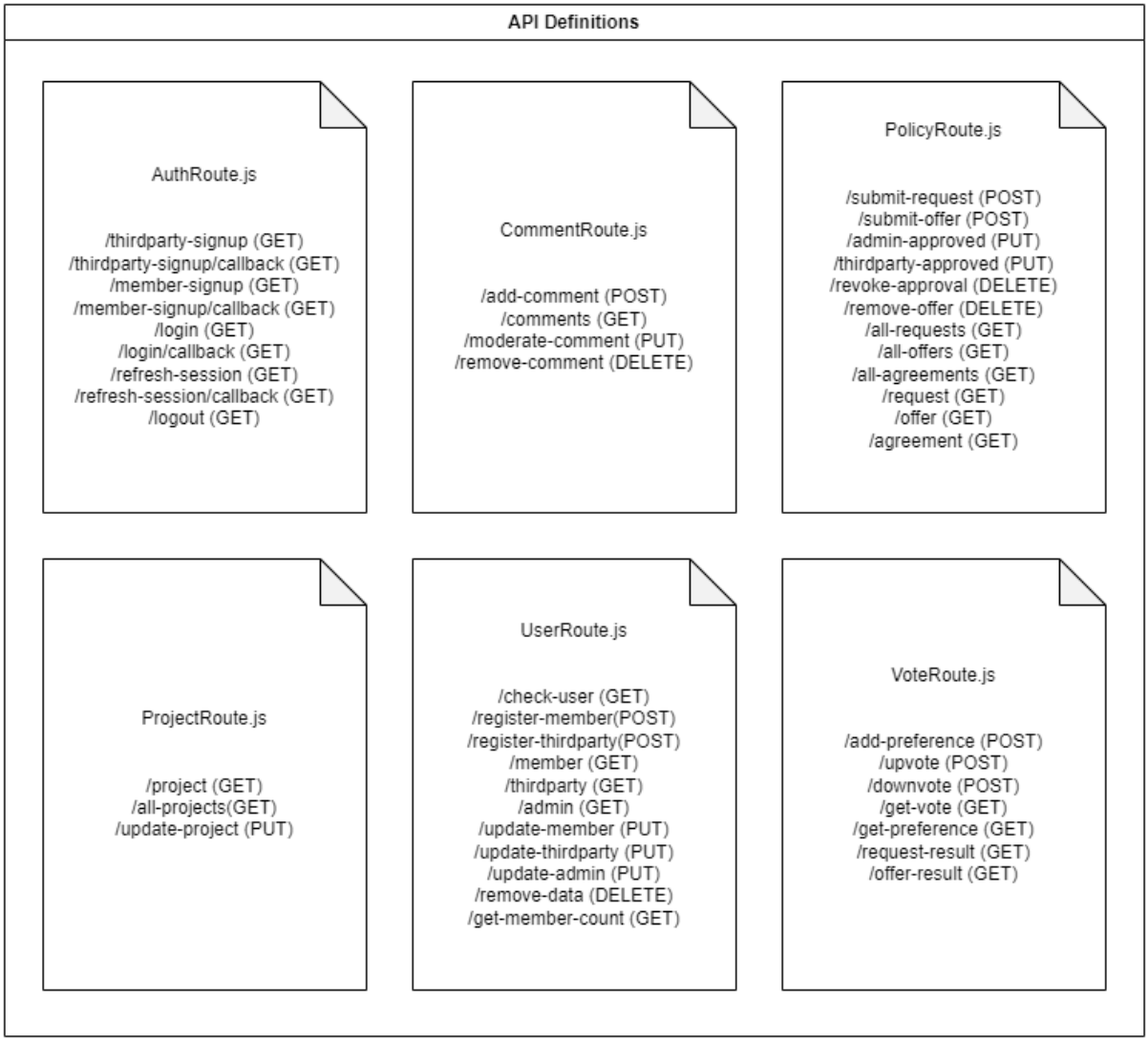


Figure C.2: API Definitions by Routes file

through the server is consistent and private.

C.4.2 Frontend Development

The pages of the web interface for DataConsensus will be examined in Section 4.4, however this section will briefly outline the structure of the frontend code to aid future developers. Within the src/app directory of the frontend code, pages related to the authentication process can be found in the Auth directory. Other pages can easily be found as components in the src/app directory. Compounds that are used repeatedly can be found in the features section, including the menu, project-status and comments.

The services directory contains the api calls divided between the comment.service.ts, policy.service.ts, user.service.ts, vote.service.ts. These services vaguely correspond to the

structure of the endpoints routes in the backend, however the policy.service.ts combined both projects and policies endpoints, while the user.service.ts includes calls for authentication. Additionally, there is also the date.service.ts, which simply provides time related functionality.

The interface files within the models directory generally match the `toJson()` format from the classes in the backend. Furthermore, the mapping.js file in the same directory contains JSONs of the different options available within the policy builders in the frontend, in the format of `[key: string]`. Using the PipeTransform module from the core Angular project, it is straightforward to convert the JSONs from a backend format, in which no spaces are allowed, to frontend, which has spaces and are more readable.

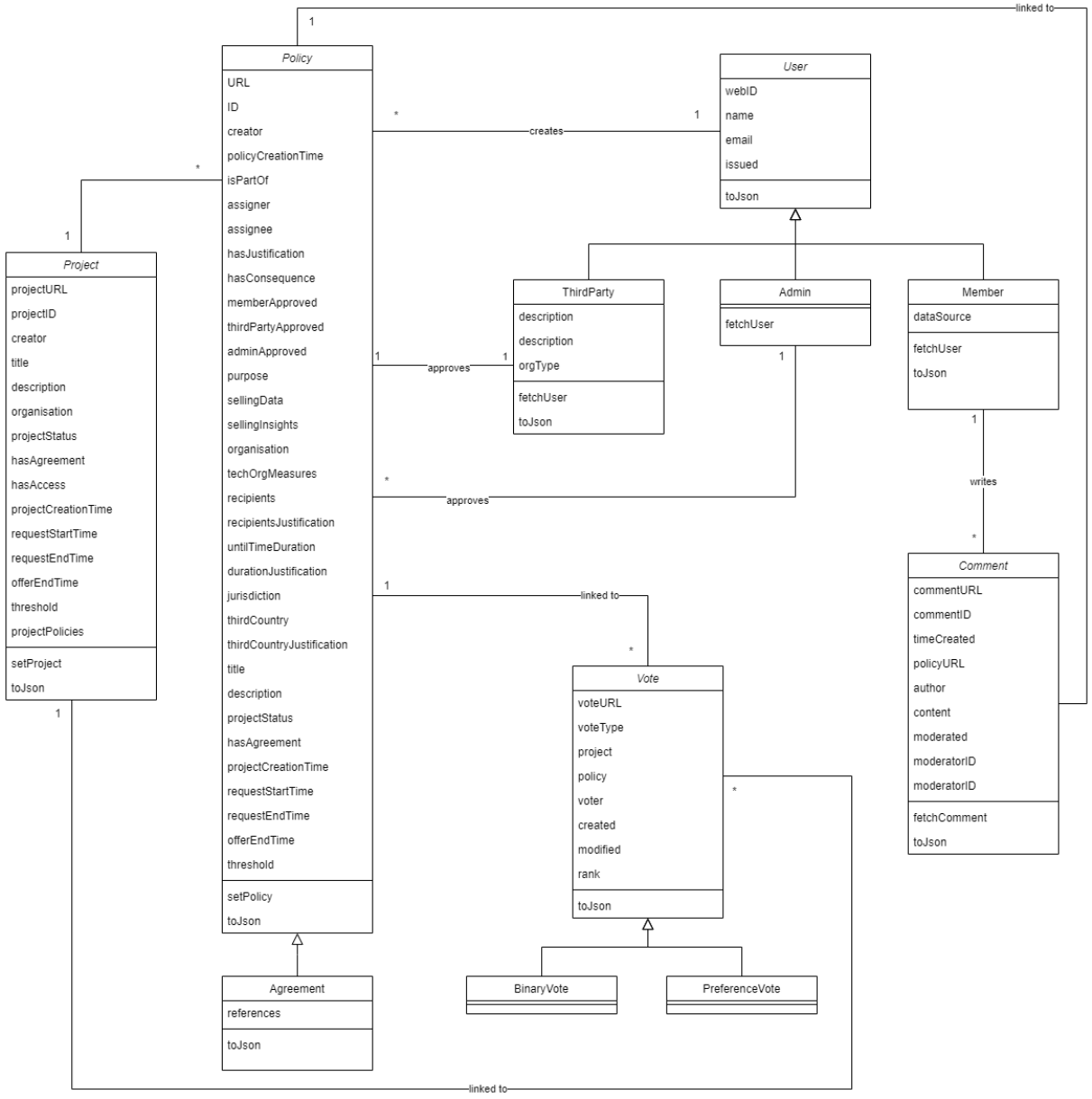


Figure C.3: Classes Diagram for the Application Layer of DataConsensus