

University of Dublin



TRINITY COLLEGE

***Data Protection Process Browser Widget:
A Visualization Tool for Interacting with Privacy Policies***

James Cox

B.A.(Mod.) Computer Science & Business

Final Year Project April 2019

Supervisor: Dave Lewis

School of Computer Science and Statistics

O'Reilly Institute, Trinity College, Dublin 2, Ireland

Declaration

I, the undersigned, declare that this work, except where otherwise stated, is entirely my own work. It has not previously been submitted as an exercise for a degree, either in Trinity College Dublin, or in any other university and that the library may lend or copy it or any part thereof on request.

A handwritten signature in black ink that reads "James Cox". The signature is written in a cursive style and is positioned above a solid blue horizontal line.

James Cox

April 2019

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Dave Lewis for providing me with support, encouragement and guidance throughout this project. I would also like to thank Harshvardhan J. Pandit for his time, feedback and invaluable opinions throughout the year.

I would like to express my gratitude to all individuals who participated in the user study and assisted in the evaluation of this project. Their willingness to contribute and their helpful feedback is greatly appreciated. I would also like to thank my family for their encouragement and continued support throughout these years in university.

Finally, I would like to express my appreciation to my colleagues, my tutor and past lecturers. Without your continued support and help with assignments, projects and countless other tasks, I would not have had the same opportunities over the past four years.

Name: James Cox

Degree: B.A. (Mod) Computer Science & Business

Project Title: Data Protection Process Browser Widget

Supervisor: Dave Lewis

Year: 2019

Abstract

Privacy policies are renowned for being difficult to read and understand due to their length and the complexity of their language. They are portrayed as being a formality, rather than something that users actually read, which means that they are often disregarded and ignored by the vast majority of consumers. As a result, they can be used in manipulative ways, allowing companies to hide or avoid certain information or practices. This project proposes a unique perspective on privacy policies which hopes to improve user interaction with these documents. While countless pages of vague, complicated text provide little value to customers prior to using a product or service, the concept of visualizing privacy policies could encourage users to make the effort to actually understand these policies, which would in turn give them a greater insight into the privacy and security of their data.

The primary goal of this project is to create a visualization of privacy policies in a manner that is personal but also insightful and beneficial. By introducing an interactive tool, organizations allow users to see first-hand where their data is being sent, what it is being used for, how it is being collected and who it is shared with. Through the tool, individuals can revoke consent, access personal data and request specific actions, thus increasing the amount of autonomy and independence that users are currently given. While the tool provides a great deal of customer value, it could also assist organizations in remaining compliant with GDPR regulation and standards.

The Privacy Policy Visualization Tool (PPVT), which is in the form of a browser add-on, can be integrated with and adapted to any company website, offering customers a formidable link between organizations and their respective privacy policies. Included within this project is a user evaluation of the visualization tool. This study was included to investigate whether the tool provides additional customer value to a number of products and services, while also analysing whether users become more informed about their data as a result.

List of Acronyms

CLIP	Centre on Law and Information Policy
CSS	Cascading Style Sheets
EEA	European Economic Area
ELI	European Legislation Indicator
FTC	Federal Trade Commission
GDPR	General Data Protection Regulation
GDPRtEXT	General Data Protection Regulation Text Extension
HTML	Hypertext Markup Language
JS	JavaScript
OPP-115	Online Privacy Policies set of 115
PIM	Privacy Indicator Metric
PPVT	Privacy Policy Visualization Tool
RGBA	Red Green Blue Alpha
SKOS	Simple Knowledge Organization System
SUS	System Usability Scale
ToSDR	Terms of Service; Didn't Read
WoT	Web of Trust

Table of Contents

Declaration	ii
Acknowledgements	iii
Abstract	iv
List of Acronyms	v
1. Introduction	1
1.1 Motivation	1
1.2 Research Objectives	1
1.3 Technical Approach	2
1.4 Research Challenges	3
1.5 Report Outline	3
2. State of the Art	5
2.1 Understanding Privacy	5
2.2 Repetition: Disregard for Privacy	6
2.2.1 Similar Research Contributions	6
2.3 GDPR: The Resurrection of Privacy?	7
2.4 Privacy By Design	8
2.5 Privacy Policies & Visualization Tools	9
2.5.1 Visualization Tool 1: Pribot & Polisis	10
2.5.2 Visualization Tool 2: Philip Raschke's GDPR-Compliant Privacy Dashboard	11
3. Design and Methodology	13
3.1 Initial Considerations	13
3.1.1 Alternative Techniques Explored	14
3.1.2 Simplification	15
3.2 System Design	16
3.2.1 Identification of Privacy Policy Aspects	17
3.2.2 Identification of Privacy Policy Issues	19
3.2.3 Layout and Structure	19
3.3 Methodology	21
3.3.1 Annotating Privacy Policies	21
3.3.2 Visualizing Privacy Policies	22
3.3.3 Interacting with Privacy Policies	23
3.4 Technologies Used	26
3.4.1 HTML & CSS	26

3.4.2 JavaScript & Vis.js	27
3.4.3 Alternative Technologies Considered	27
3.5 Use Cases and Visualization Examples	28
4. Implementation and System Overview	30
4.1 Annotation to Visualization Decisions	30
4.2 Network Physics Opportunities	31
4.2.1 BarnesHut Model	31
4.2.2 AvoidOverlap Constant	32
4.2.3 SpringLength & Damping Constants	32
4.2.4 Drawing Issues and Seed Allocation Constants	32
5. Evaluation	34
5.1 User Evaluation	34
5.1.1 Consent	34
5.1.2 Visualization Questions	35
5.1.3 Systems Usability Scale	38
5.1.4 Ethical Considerations	40
5.2 Privacy Evaluation	40
5.2.1 The States of Privacy	41
5.2.2 Human Factors	41
5.2.3 Data Factors	42
5.2.4 Service Factors	43
5.3 Evaluation Summary & Limitations	45
6. Conclusion	46
6.1 Objective Assessment	46
6.2 Project Issues	46
6.3 Future Work	47
Bibliography	49
Appendices	52

List of Tables

<i>Table 2.5 Comparison of Privacy-Enhancing / Awareness Tools (Khajuria et al., 2017).....</i>	<i>10</i>
<i>Table 3.3 1 PPVT Aspects and Respective Colours.</i>	<i>21</i>
<i>Table 3.3 2 PPVT Aspects and Interaction Possibilities</i>	<i>25</i>
<i>Table 5.1.2 Visualization Questions: Correct Answer Percentages.....</i>	<i>38</i>
<i>Table 5.2.1 The States of Privacy (Kosa, El-Khatib and Marsh, 2011).....</i>	<i>41</i>
<i>Table 5.2.2 1 An Adaptation of the Human Factors Set (Kosa et al., 2011)</i>	<i>42</i>
<i>Table 5.2.2 2 Evaluation Table: Human Factors</i>	<i>42</i>
<i>Table 5.2.3 1 An Adaptation of the Data Factors Set (Kosa et al., 2011).....</i>	<i>43</i>
<i>Table 5.2.3 2 Evaluation Table: Data Factors.....</i>	<i>43</i>
<i>Table 5.2.4 1 An Adaptation of the Service Factors Set (Kosa et al., 2011).....</i>	<i>44</i>
<i>Table 5.2.4 2 Evaluation Table: Service Factors</i>	<i>44</i>
<i>Table 5.2.4 3 Evaluation Table: Privacy State.....</i>	<i>44</i>

List of Figures

Fig 2.1 Purpose of Privacy Policies (P. S. Dhotre et al., 2016).....	5
Fig 2.2.1 Summary of Privacy Issues (Centre of Law and Information Policy, 2014)	7
Fig 2.5.1 1 A high-level overview of Polisis (Harkous, Fawaz, Lebret, Schaub, Shin & Aberer, 2018).....	11
Fig 2.5.1 2 An example of an interaction with the Polisis Visualization Tool regarding the BMW Privacy Policy.....	11
Fig 2.5.2 The Standard View of Raschke’s Privacy Dashboard.	12
Fig 3.1.1 1 Compliance Fault Tree designed using the DPL Fault Tree System. Copyright © 2018 Syncopation Software, Inc.	14
Fig 3.1.1 2 Compliance Circuit Diagram designed using the DPL Fault Tree System. Copyright © 2018 Syncopation Software, Inc.	15
Fig 3.1.2 An example of the Motif Simplification being put into practise for a Network Diagram.....	16
Fig 3.2 System Design of the PPVT.....	17
Fig 3.2.1 Examples of the Data Sharing and Legal Basis Aspects of the Just Eat Policy along with their respective colours.	18
Fig 3.2.2 Examples of some of the Issues included in the Just Eat Privacy Policy	19
Fig 3.2.3 The relationship between Issues and the Data Sharing Aspect in the Data Sharing Tab of the Just Eat Policy.	20
Fig 3.3.1 A Section of the Just Eat Policy which has been annotated using the pre-determined aspects and their respective colours.	22
Fig 3.3.2 A picture taken from the Data Retention tab. By hovering over the rectangular issue, other related aspects can be seen.....	23
Fig 3.3.3 An example of how the Rights Tab of the JustEat PPVT offers users with much more interaction and possibilities in comparison with standard policies or organization websites.....	26

Fig 3.4.3 The Designer-Reader-Data Trinity (Illiinsky & Steele, 2011)	28
Fig 4.1 The Decision Tree which was used when choosing the correct information to include.	30
Fig 4.2.1 An example of some of the constants and customization possible with the BarnesHut model.....	31
Fig 5.1 The Introduction Page of the User Evaluation	34
Fig 5.1.1 1 The Consent Form Section of the Evaluation	35
Fig 5.1.1 2 A Pie Chart representing Users and Consent for the Evaluation.....	35
Fig 5.1.2 1 Question 1 of the Visualization Survey.....	36
Fig 5.1.2 2 Response to Question 1.....	36
Fig 5.1.2 3 Question 3 of the Visualization Survey.....	36
Fig 5.1.2 4 Response to Question 3.....	37
Fig 5.1.2 5 Question 5 of the Visualization Survey.....	37
Fig 5.1.2 6 Response to Question 5.....	37
Fig 5.1.3 1 Question 1 of the SUS.....	38
Fig 5.1.3 2 Question 4 of the SUS.....	39
Fig 5.1.3 3 Question 10 of the SUS.....	39
Fig 6.2 Gantt Chart introduced during this project.	47

1. Introduction

1.1 Motivation

In recent years, there has been major concern regarding the decline of online privacy. Countless security breaches along with scandals (Verizon, 2018) and accusations of consumer manipulation have led to a lack of trust between the public and certain organizations. When uncertainty arises, individuals become less likely to accept many products and services (Solin, 2017). To counter-act this issue, organizations have attempted to make their products more transparent and trustworthy. They put monumental effort into how services are designed, while also investing time and money into how their respective privacy policies are portrayed (OECD, 2010). The problem that many still ignore is that users simply refuse to study these privacy policies.

There are a number of reasons as to why the public avoid reading privacy policies. In a paper written in 2008 about the cost of privacy policies, it is mentioned that if all Americans were to read word-for-word the terms of service of each site or product that they accessed, the accumulated time spent by the nation would amount to 54 billion hours reading privacy policies (McDonald & Cranor, 2008). This is the equivalent of 40 minutes per day for each person, proving that people simply do not have the time to read these policies. In addition to time constraints, another aspect is the fact that the majority of the public either have difficulty in understanding the language of privacy policies or lack the engagement and focus to concentrate on a document that may take 10 – 15 minutes to read. Attention is one of the most powerful assets of human beings, but concentration levels have decreased in recent years (Lamba et al., 2014). According to Hartley and Davis, the attention span of the average adult on a specific subject is approximately 10 minutes at a time. While a lapse in concentration may only last for a moment, it still reiterates the fact that anything over 10 minutes of apparently unnecessary reading becomes difficult for the average user. An ideal alternative solution would be to make these policies more visual and more interactive. Humans relish the opportunity to get involved with products that contain interactive activities, as it can make certain tasks seem more interesting and inclusive (NIHRC, 2008). Reading privacy policies remains a tedious and boring task, but the possibility of visualizing them and integrating a level of interaction holds the potential to create a new level of interest and understanding for the benefit of both users and service providers.

1.2 Research Objectives

The way in which privacy policies are designed and presented has constantly been criticized (Marotta-Wurgler, 2015). They are usually too long and too tedious to fully understand. This project looks specifically at the following question:

“Can users become more informed about their data if privacy policies are presented in a visual and interactive way rather than in the standard textual format?”

To address and explore the above research question, certain objectives must be outlined:

- Establish a visual design that improves the user’s understanding of privacy policies while still displaying all necessary information.
- Increase emphasis on consent and rights, particularly regarding GDPR changes, compared to other available tools and services.
- Conduct a user evaluation to measure the effectiveness of the tool and also to gain an insight into user opinion of the visualization.

When beginning this thesis, agreeing upon the above objectives was vital to ensuring that the research question was meaningfully explored. While the primary goal of the project was to visualize and shorten privacy policies, it was important to be reminded that privacy policies still hold information that must be present. Finding the right mix of keeping users interested and covering all of the essentials was something that was recognised at an early stage. In addition, one of the initial objectives was to provide a detailed focus on aspects of privacy policies such as consent and rights. These aspects have been under the spotlight in recent years, with scandals and malpractices such as Facebook’s unlikely relationship with Cambridge Analytica being of particular interest. By having a greater emphasis on these aspects, users can see first-hand what services they have given consent to and how changing this consent can affect other aspects of the privacy policy.

It was decided that the most effective way to measure the success of the visualization tool was to perform a user evaluation. By inviting the public to interact with the software, results and opinions can be recorded which show what parts of the tool provide value and which ones don’t. The study also proves whether the tool is informative as well as easy to use. Asking logical questions and allowing the user to interact with the tool to find the answer is a very valuable way of determining the success of a product.

1.3 Technical Approach

To satisfy the above research objectives, it was decided that a visualization tool was to be designed that assisted users in understanding how their data is collected, stored, processed and shared, while also ensuring that organizations adhere to GDPR standards and prioritise the protection of customers. The Privacy Policy Visualization Tool (PPVT) was developed during this project. The PPVT essentially involves a process of replacing standard text-based policies with

visualizations that simplify policies and highlight important information through a process known as annotation. Visualizations work in coherence with standard policies to improve user insight into the management of their data and interactive features offer users an alternative method of learning about policies and data.

1.4 Research Challenges

During the course of this thesis there were certain challenges which had to be overcome in order to produce a successful visualization tool. Initially there was the issue of subject knowledge. Privacy policies and GDPR are both quite specific topics which required improved understanding and research. In addition, there are few similar projects or tools available regarding these topics as a collective. While this reiterates the demand for this type of tool, it was difficult to initially gain any insight into what a final product may look like. The most time-consuming process of the project has been annotating the standard privacy policies. Reading through entire policies and categorizing the different issues and aspects was a very tedious task. However, once the policies were annotated, the rest of the project became easier and the process also improved the researcher's knowledge of the subjects at hand. When deciding upon the method of evaluation, it became evident that a successful ethics clearance application was required. While clearly important in ensuring that the user study was both correctly designed and provided individuals with the correct consent and information sheets, the process was quite tedious and time consuming.

1.5 Report Outline

This report describes and documents the research undertaken, the design, the technology implemented and the evaluation that was executed over the duration of this project. These decisions, along with their possible alternatives, will be discussed in detail below.

Chapter 2 provides an overview of the state of the art along with some of the background research regarding privacy policies and the need for change. Readings were recommended with a particular focus on the feasibility and possibility of introducing GDPR-compliant privacy dashboards. Alternative concepts included designing a machine learning based policy reader and a GDPR compliance interactive circuit.

Chapter 3 outlines the design of the system, along with the technologies used during the project. The key decisions and methodology are also examined further here. Why certain

technologies and libraries were chosen over others is something that took a great deal of consideration at the beginning of this thesis.

Chapter 4 follows up from the design aspects of chapter 3, with more of a focus on the implementation details of the visualization tool. The tool is also contrasted with possible alternatives and similar products available to users. A system overview is included, which demonstrates the complete functionality of the system, giving the reader a more detailed perspective of the tool.

Chapter 5 focuses mainly on the follow-up analysis, particularly regarding the user study and evaluation which began towards the end of the project. The data collected from the online questionnaire is analysed and results and conclusions are drawn up, while future work and other considerations are also mentioned.

Chapter 6 concludes the report with final thoughts on whether all objectives and research questions were recognised and dealt with successfully.

2. State of the Art

2.1 Understanding Privacy

“To be left alone is the most precious thing one can ask of the modern world” (Burgess, 1986). The concept of privacy is no doubt a social concept (Lou & Ren, 2009). It is a luxury that seems to have been forgotten or ignored in this modern era. As a result of rapid digitization and constant technical advances over the past few years, it has become very difficult to define the boundaries of an individual’s privacy (European Parliament Policy Department, 2015). Trends and new technologies such as big data, machine-learning and artificial intelligence offer service providers with countless new avenues to explore regarding data processing, which in turn leaves service users increasingly more vulnerable. In truth, while the definition of privacy has changed over time (Lukács, 2016), the public’s expectations and demands regarding privacy remain the same.

Privacy policies are essentially legally required documents that explain exactly how a company gathers, uses, processes and manages a customer or client’s personal data. In addition, laws such as the CalOPPA in California and more recently, the GDPR in the EU have forced organizations to reconsider the ways in which they represent customers and display privacy policies (Kurtz et al., 2018). While mandatory for all products and services, they are also expected to inform and educate users about how their privacy and their personal information is protected. However, this is not always the case. Organizations include privacy policies for a number of different reasons. Below is the result of a survey introduced by P.S. Dhotre that conveys the main reasons as to why, other than for legal reasons, a company produces a privacy policy:

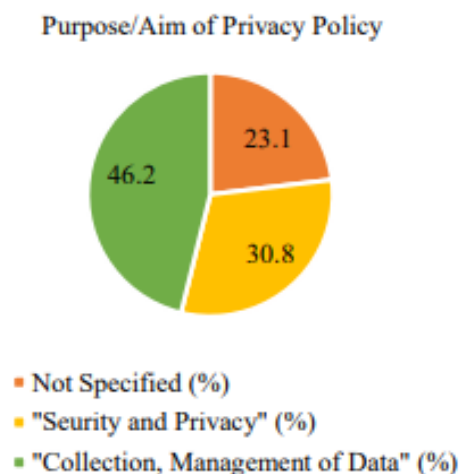


Fig 2.1 Purpose of Privacy Policies (P. S. Dhotre et al., 2016)

A privacy policy should focus on the protection of personal information. However, the majority of policies (46.2%) prioritise informing users of how data is collected and managed, while 23.1% of policies do not specify the purpose of the document. The official purpose of privacy policies according to GDPR standards is to give individuals control over their personal information (Council of the European Union, 2015), and the above survey proves that organizations remain unsure of this.

2.2 Repetition: Disregard for Privacy

When considering the extent to which user information is exploited and manipulated, countless cases come to mind. Regardless of how much digitization and expansion affect privacy, organizations have always had the tendency to disregard privacy in some alternative shape or form (Pelteret & Ophoff, 2016). Whether it is the recent scandal involving Cambridge Analytica manipulating Facebook users (Maltby, 2018) or the Sony CD Spyware incident (Halderman & Felten, 2006), if regulation on data protection remains unclear, profitability and corporate success will continue to be prioritised over privacy and customer safety. In addition to this, security breaches and increasingly sophisticated viruses and malware have increased the level of distrust that individuals hold towards certain organizations.

2.2.1 Similar Research Contributions

There have been several research contributions aimed at understanding users' opinions and outlooks on privacy. In an attempt to raise awareness and investigate online practices of information handling and related privacy issues, a survey analysing 116 complaints was carried by the Federal Trade Commission (FTC) (CLIP, 2014). The survey presented the list of petitions filed by individuals and their classifications. It illustrates four main categories of privacy issues; Unauthorized Disclosure of Personal Information, a Surreptitious Collection of Data, Inadequate Practices and Wrongful Retention of Personal Information. The complaints consist of issues including personal information sharing, no consent, unauthorized access, no notification, extensive information collection, inefficient information protection system, incomplete privacy policies, and retention of information for a longer time than necessary.

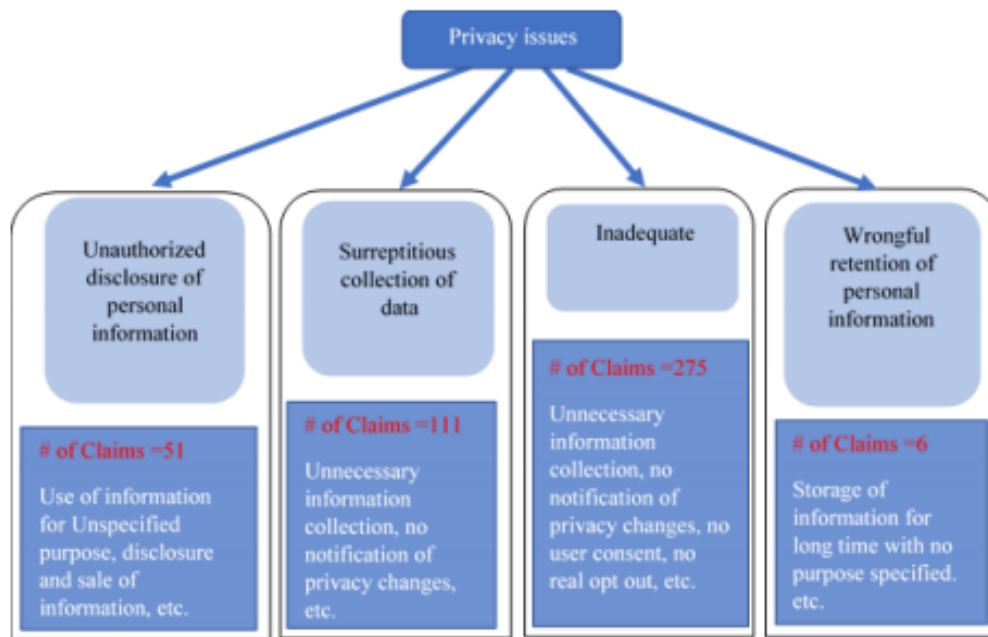


Fig 2.2.1 Summary of Privacy Issues (Centre of Law and Information Policy, 2014)

While the above survey shows varying issues regarding privacy and online malpractices, the claims as a collective resonate some of the vague standards and expectations associated with user privacy. The question that must be asked is whether an informative and interactive privacy tool would have led to the non-existence of these claims? A tool of this nature would both inform users of how data is retained and processed while also ensuring that organizations don't hide certain important information in the fine print of policies.

2.3 GDPR: The Resurrection of Privacy?

Data has become an integral part of the every-day lives of consumers, especially regarding privacy and security. The introduction by the EU of the General Data Protection Regulation (GDPR) on the 25th May 2018 has heightened the obligation for service providers to obtain clear, informed consent from individuals for the use of personal data. It also greatly increases the regulatory penalties for non-compliance when handling that personal data. This new regulation ensures that all user information is treated with the utmost care, wherein organisations are required to provide information about how they collect and use data, along with placing more consideration on consent and its consequences. GDPR also provides the data subject with various rights. These include:

- The Right to be Informed
- The Right of Access
- The Right to Rectification

- The Right to Erasure
- The Right to Restrict Processing
- The Right to Data Portability
- The Right to Object
- Rights in Relation to Automated Decision Making and Profiling

At the moment, there are existing and ongoing efforts to assist users in understanding privacy policies and exercising their rights through summaries or visualizations. According to their first annual report, the number of data breaches reported to the Data Protection Commission increased by 70% in 2018, primarily as a result of the introduction of the GDPR (Data Protection Commission, 2018). In addition to this, the volume of complaints rose by over 50%, with the largest amount of complaints being related to the right of access to personal data. This increase in complaints shows that users are now more aware of their rights regarding how their information is processed and new products and services could ensure that service providers guarantee compliance with new standards and regulation.

2.4 Privacy By Design

While the introduction of the GDPR has forced organizations to become more considerate about users and their respective data, there still exists a need for products and services to further educate and inform individuals about their personal information and how it is treated. One possible consideration when creating these types of products is to focus on the privacy risks during the design phase. Privacy by Design is a framework that is used during the design phase of a product or service which places emphasis on privacy by default (Cavoukian, 2012). The framework promotes an approach to design that is proactive rather than reactive. By anticipating possible threats before they occur, many possible breaches or other worrying alternatives can be avoided. Also, by implementing privacy as the default, users are not required to make any initial changes to services when considering the protection of their information. The model also focuses largely on transparency and visibility, which would greatly improve the current trust levels between entities in a new product or service. During the design stage of the Privacy Policy Visualization Tool (PPVT), Privacy by Design became an integral reference to use when acknowledging user information and the way forward in data protection.

2.5 Privacy Policies & Visualization Tools

User information is gathered by service providers at a high rate and at large scale. As a result of the technical advances in today's digital world, users have a lot of convenient services that make their daily tasks easier. Multiple visualization tools and interactive dashboards have been designed with the intention of assisting users with options such as the opportunity to avoid being tracked online or blocking third-party data processing. Below is a table based on the work and typology of Khajuria that represents some of the readily available visualization tools, along with a description of each:

<i>Tool Name</i>	<i>Functionality</i>	<i>Type of Tool</i>
Privacy Badger	This tool helps users to block tracking from advertisers and third parties.	Blocking
Lightbeam	Using this tool, the user will be aware of how the first and third-party websites interact and their relationships.	Awareness
Disconnect	Unsecured connections and hidden requests for users' personal information are visualized by this tool. Also, this tool allows the user to block trackers and hackers.	Awareness & Blocking
Ghostery	Detecting and blocking of invisible trackers is the main objective of this tool.	Blocking
MyPermission	This tool gives users complete control over those apps that access the users' data.	Control & Blocking
Terms of Service; Didn't Read (ToSDR)	Using ToSDR, rating, and labelling of the terms and privacy policies of major websites can be seen, based on a user community. The ratings cover a range from very good (Class A) to very bad (Class E).	Privacy Awareness

Web of Trust (WoT)	Rating of websites based on user comments/community.	Trust Awareness
--------------------	--	-----------------

Table 2.5 Comparison of Privacy-Enhancing / Awareness Tools (Khajuria et al., 2017)

While the above tools are incredibly useful, both raising awareness and resolving trust issues, none of these tools provide an overall solution to issues of blocking, awareness and interaction. With the above tools, the price of these services is the loss of privacy, as many of these companies are accessing more than necessary amounts of personal information (Stephenson, 2015). Individuals still need more insight into how their information is processed, collected and shared, along with what exact information is being processed (Waldman, 2018). Privacy policies that service providers release, while improving because of new regulation, are still complicated and difficult to understand, and hence the need for a visualization tool that allows users to fully understand what is happening to their data without having to read countless pages of text is very much in demand.

2.5.1 Visualization Tool 1: Pribot & Polisis

Privacy policies are essentially an organizations primary channel for informing users about data collection and sharing practices. As mentioned previously, these text-based policies are long and tedious for an individual to analyse. Users still lack usable tools that assist with the depth and complexity of privacy policies. As a result, in 2018 the developers behind pribot.org released a ground-breaking product known as Pribot. The initial intention for Pribot was that users could ask this bot questions about a certain privacy policy and the bot, through the help of artificial intelligence, would provide the answer based on what is mentioned in the privacy policy text. The service, which took the form of a conversation with a chatbot, replies to questions with snippets of text from policies, ensuring that users don't have to manually search for a particular query.

While developing the Pribot, the opportunity arose to introduce a by-product, which is called Polisis. The primary objective of the Polisis system is to automatically analyse privacy policies using machine learning and to visualize the results in an informative and interactive way. The tool allows users to gain an insight into the information that is collected, the data that is shared with third parties, what security measures are implemented and many other aspects of privacy policies. While initially aimed at the users of said products or services, Polisis and Pribot have also become beneficial for regulators when assessing privacy policies to ensure that organizations remain in compliance with GDPR standards. The true success of Polisis originates from the way in which it was designed compared to other visualization tools. There are no readily available public datasets to assist with the machine learning techniques, so a new approach had to be undertaken.

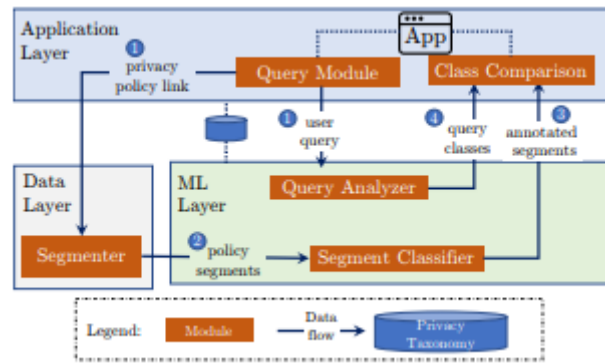


Fig 2.5.1 1 A high-level overview of Polisis (Harkous, Fawaz, Lebre, Schaub, Shin & Aberer, 2018).

Polisis was instead designed by working backwards, where the priority was labelling certain segments of each policy and then implementing solutions to match those segments to specific questions using machine learning techniques (Harkous, Fawaz, Lebre, Schaub, Shin & Aberer, 2018). This was managed by initially training a word embedding model on a number of privacy policies and then training a hierarchy of 22 classifiers for labelling the different aspects of each policy. The OPP-115 dataset was referenced for this process, which was created for the Usable Privacy Project.



Fig 2.5.1 2 An example of an interaction with the Polisis Visualization Tool regarding the BMW Privacy Policy.

2.5.2 Visualization Tool 2: Philip Raschke's GDPR-Compliant Privacy Dashboard

The role of personal data has gained significance across all business domains in recent years. Despite the strict legal restrictions that processing personal data is subject to, users tend to respond to the extensive collection of data by service providers with distrust. Philip Raschke decided that a GDPR-compliant privacy dashboard could be the solution to improving trust levels between companies and their users. His version of a privacy dashboard aims to provide a means of managing GDPR rights when availing of a specific product or service. Using technologies such as JavaScript and

React Native, Raschke provides a supportive tool which offers the informative nature of a privacy policy but with more of an interactive perspective.

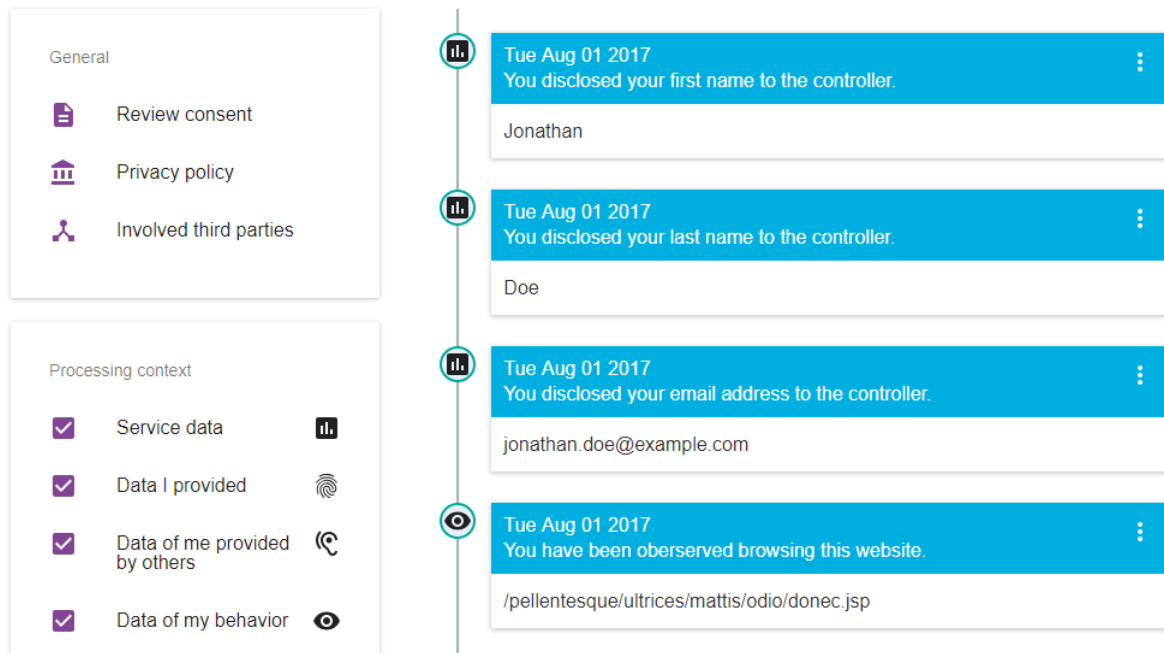


Fig 2.5.2 The Standard View of Raschke's Privacy Dashboard.

The dashboard is similar in structure to the Polisis service regarding informative language and detailed descriptions in the privacy policy and third-party tabs. However, it seems to offer more in the consent and rights aspects. The opportunity to access all consent agreements along with being able to filter the types of information collected and processed ensures that users gain a valuable insight into how their information is treated by organizations.

Raschke's architecture and initial idea for the dashboard is the ideal mix of simplistic design and sophisticated implementation. While complex in its back-end development, the project acknowledges the problem that the dashboard is resolving. Users need a clear design, especially regarding visualizations. Raschke designs an aesthetic and informative dashboard which allows users to immediately understand the message being conveyed by organizations.

3. Design and Methodology

3.1 Initial Considerations

When designing a product or service such as a visualization tool, it is important to first note the reasons as to why the tool should be created. By acknowledging the flaws of previous products, along with why users are so dissatisfied with text-based privacy policies, it became much easier to design a visualization tool that provides value and learns from the mistakes of other attempts. The four most important aspects regarding current privacy policy ineffectiveness are:

- User awareness and misunderstanding of privacy
- User acknowledgement of privacy policies and their purpose
- The lack of user control with consent in particular
- How new standards such as the GDPR affect user rights

By including these considerations in the design phase of the development cycle, the visualization tool essentially takes each issue into account. Regarding the awareness and misunderstanding of privacy, informative tabs are available which educate the user on how they should benefit from privacy along with insight into certain technological concepts such as cookies and how information is collected from a user's device. To inform users of the purpose of privacy policies, detailed descriptions of the aims of each privacy policy are included. In addition, the visualizations are structured in such a way as to convey the message of how all aspects and issues in privacy policies are related through the common processes and transferred data. The lack of interaction with user consent is an issue that has not been addressed enough in previous visualization tools. This was one of the first issues taken into consideration, and a consent tab was included in the design as a result. This tab allows users to revoke consent from certain services, while also informing the user on what these consent changes will mean for other aspects of the relationship between users and the rest of the services. Along with consent, the other primary consideration was how the tool would account for GDPR standards. Since the regulatory changes have only been introduced since May 2018, few interactive services prioritise the standards other than the tools mentioned above. This visualization tool allows users to become informed on their new rights, along with offering them the opportunity to avail of some of these rights, including The Right to Data Portability and The Right to Erasure. This GDPR focused tab also enables users to opt-out of certain services and to unsubscribe from marketing communications, which many users have had issues with in the past.

3.1.1 Alternative Techniques Explored

After taking initial considerations into account, the next developmental step was to decide upon the type of design that was to be created. This included the techniques that should be used and there were a number of options. The very first technique that was considered was to structure the visualization with a completely different perspective than other tools. Initially, the project was to focus more on compliance and how companies could ensure that they remained in line with GDPR standards by protecting and respecting the privacy of user information. This idea led to an original design that took the shape of a circuit. Each node on the circuit represented a barrier that needed to be passed to remain in compliance with regulation. Each node would be an aspect or concept associated with GDPR and the optimum privacy policy. Examples of nodes included:

- An option for users to revoke consent
- Whether the privacy policy mentioned how long data was retained
- If contact details were given for each third-party service provider
- An option for users to practice a new GDPR right. Eg: The Right to Erasure

If the policy complied with standards regarding a certain aspect, the node would be green.

Otherwise, the node would turn red. If the visualization created a path from beginning to end where all nodes were green, the node at the end would turn yellow, approving the policy and hence creating a circuit-like diagram.

The idea for this technique was established through the DPL Fault Tree, developed by Syncopation Software (Silvianita, 2013). The tree design, while similar in concept, did not focus on privacy policies, but on the GDPR articles and whether they specifically were acknowledged among organizations. The circuit diagrams were also presented alongside truth tables and probability charts, which were also analysed and considered during the early stages of this project. Below is an example of the type of diagrams that were designed using DPL:

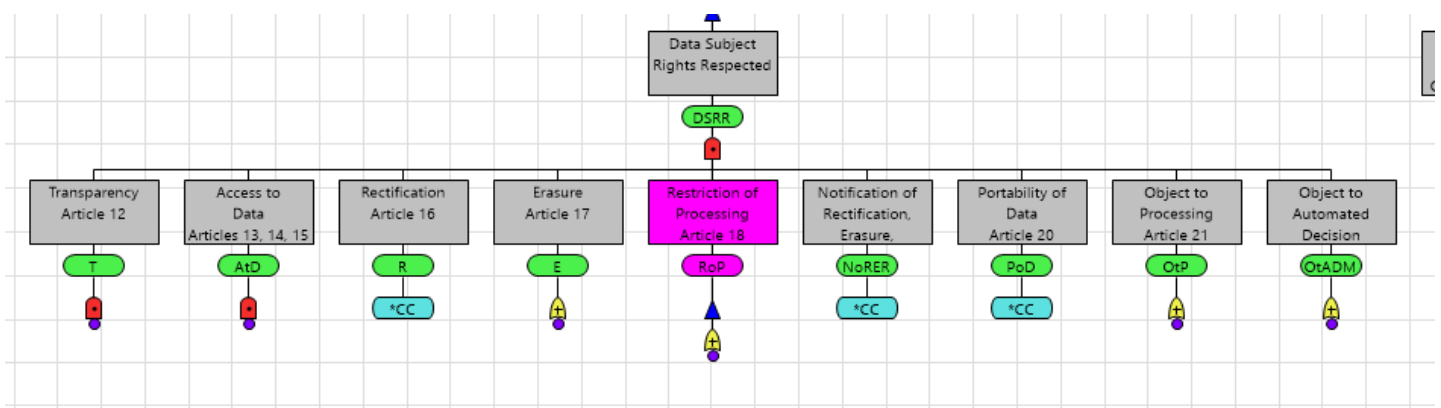


Fig 3.1.1 1 Compliance Fault Tree designed using the DPL Fault Tree System. Copyright © 2018 Syncopation Software, Inc.

The expandable nodes contain detailed information and the relationships between the different aspects which would have been of considerable use if this approach continued. Below is an overall view of the circuit which would have shown whether a policy was in compliance or not:

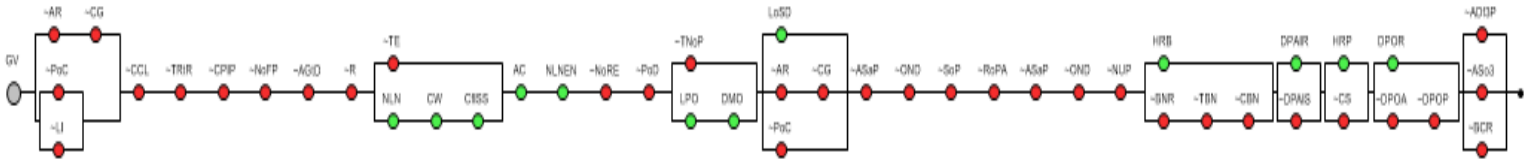


Fig 3.1.1 2 Compliance Circuit Diagram designed using the DPL Fault Tree System. Copyright © 2018 Syncopation Software, Inc.

While initially this technique seemed to provide value to companies and service providers, it seems to offer more to these entities than to the users and individuals, who should be the priority when considering the protection of user information. In addition, the DPL system offered the opportunity to interact with the nodes, but it didn't allow for the revoking of consent or the updating of permission regarding marketing communications or data portability. One of the initial objectives for this project was to provide greater focus on the needs of the data subject rather than the needs of the data controller, hence alternative options were considered.

3.1.2 Simplification

One of the main reasons for certain previous projects failing to design an informative and interactive privacy policy tool is that the majority of the products were actually overly complicated. The reason that there is demand for a tool of this nature is because of the complexity and length of standard textual policies. Visualization tools should prioritise the simplification of these policies with a primary goal of informing users of the protection of their personal data in a shorter period of time than it would take to come to the same realisation using the original policies. While tools such as Polisis and the Privacy Dashboard provide valuable insight for users regarding how their personal information is processed, at times these visualizations appear slightly over-complicated. As a result of the type of overlapping and interlinking relationships between aspects and issues in most privacy policies, it was decided that a network diagram would be implemented during visualization. When it was decided that the PPVT should take the form of a network diagram, the faults of previous visualizations were constantly considered. One technique which became increasingly useful during this project was Motif Simplification. This process involves the replacement of common patterns and nodes with compact and meaningful glyphs and was designed by Cody Dunne and Ben Shneiderman in an attempt to improve network visualization readability (Dunne & Shneiderman, 2013). The

benefits of utilising a technique such as Motif Simplification include the fact that nodes then require less screen space and the visualization tends to provide the optimum amount of information about each aspect.

Motif Simplification became useful during the initial design phase of the PPVT when the plan was for nodes to exist for each reference of an issue or aspect. The network diagram resultingly became overcrowded and meaningful comparisons and analysis were impossible. The visualization was simplified by increasing the size of nodes based on the amount of times that certain aspects are mentioned in the policies, which led to a much more effective display and analysis. Documenting one instance of each aspect or issue and providing in-depth details upon further interaction from users improved the diagrams readability and increased the engagement levels of individuals, which is evident in the evaluation section. For topics that tend to come under criticism for complexity and length, it was important to simplify any visualizations that dealt with privacy policies and the GDPR in particular.

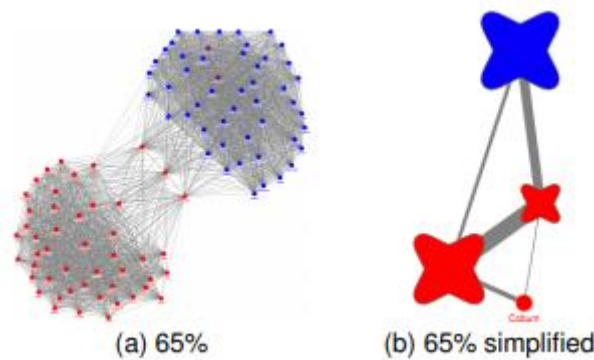


Fig 3.1.2 An example of the Motif Simplification being put into practise for a Network Diagram (Dunne & Shneiderman, 2013).

3.2 System Design

As a result of the demand and necessity for a new way of understanding and analysing privacy policies, it was decided that the Privacy Policy Visualization Tool (PPVT) should be designed. The PPVT is an alternative method for interacting with policies to ensure that users are fully aware and confident of the protection of their personal information as well as how their information is collected, stored, processed and shared. Some of the main ideas for the design included:

- A Network Diagram so that the relationship between different aspects and issues could clearly be conveyed.

- A simple, yet effective layout. White was chosen as the primary colour and different colours were assigned to each aspect of a policy. White ensures simplicity and avoids confusion between colours clashing.
- Acknowledge the fact that the tool will be a browser add-on. The PPVT should be considered a complimentary tool that works in coherence with websites and standard policies. Users still have access to company websites and the standard text-based policies should also be available for people who prefer a text-based structure.

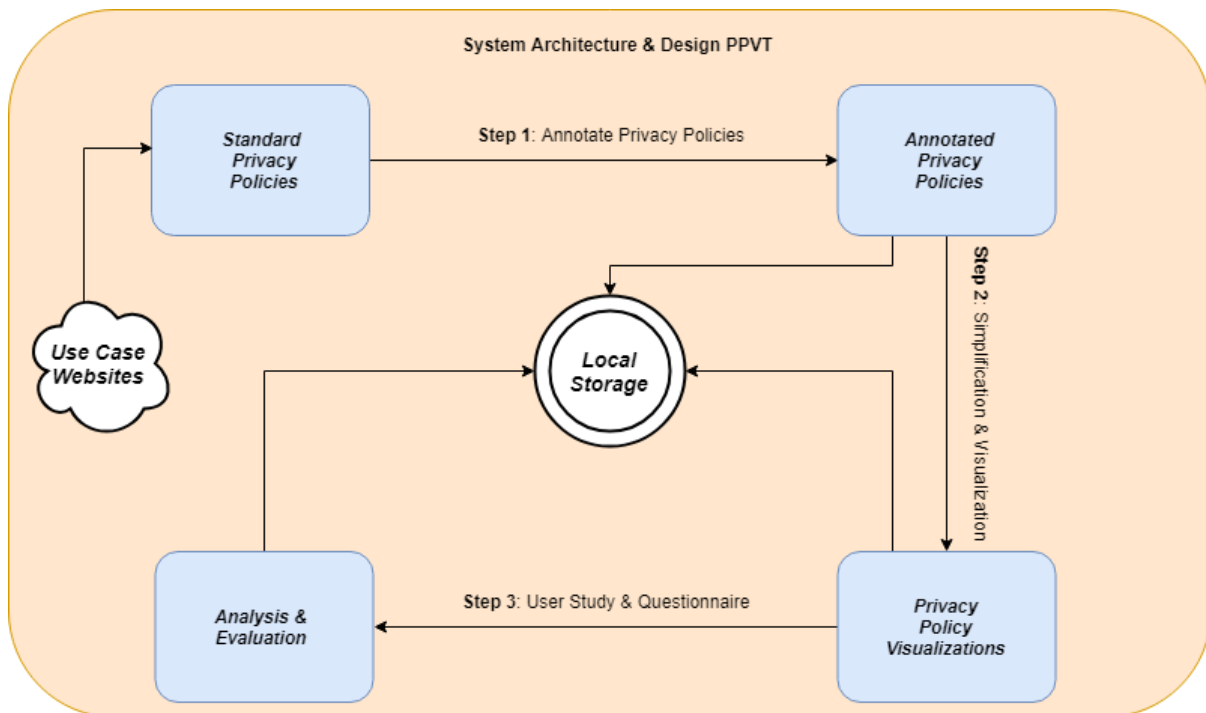


Fig 3.2 System Design of the PPVT.

The above design portrays a process where standard policies are taken from product websites and annotated manually using HTML. These annotated policies are then converted to visualizations, adapting a mode of simplification and privacy by design. The information is displayed in a network diagram to accurately portray the relationships of nodes in these policies. The tool is then evaluated using the user study and the researcher's evaluation, and all relevant information is stored locally.

3.2.1 Identification of Privacy Policy Aspects

The first step when designing the PPVT was to establish a way of differentiating the numerous aspects of a privacy policy. In a publication released in 2018, Pandit, Fatema, O'Sullivan and Lewis investigate GDPR as a linked data resource. Their GDPR text extensions (GDPRtEXT) uses the European Legislation Identifier (ELI) ontology along with Simple Knowledge Organization System

(SKOS) vocabulary to represent the GDPR as linked data (Pandit, Fatema, O’Sullivan & Lewis, 2018).

The PPVT references this work when identifying the aspects that make up privacy policies. These include:

- Data Type – Eg: Name, Age, Address
- Data Category – Eg: Personal Information, Delivery Details
- Legal Basis – Eg: Contractual Reasons, Legitimate Business Interest
- Third Party – Eg: Delivery Partners, Subsidiaries & Affiliates
- Data Sharing – Eg: Law Enforcement & Regulatory Bodies
- Consent – Eg: Process Personal Information, Share Geo-Location
- Rights – Eg: Opt-out of Marketing Communications, Revoke Consent
- Data Retention – Eg: Upon Account Deletion, 90 Days
- Automation – Eg: Cookies & Other Technologies
- Process – Eg: Marketing Services, Fraud Prevention
- Location – Eg: Countries within the EEA
- Processor – Eg: Service Providers, Payment Processors
- Data Source – Eg: Provided by User, Provided by Third Party

The above “Aspects” are the key link between different issues such as who information is shared with or your data protection rights. Certain issues might be associated with multiple aspects. For example, who information is shared with can be linked to the third-party aspect while legal basis is obviously a related factor too. In addition to this, each aspect was assigned a colour and these colours are associated with the respective aspects throughout the visualization, increasing familiarity and customer value. By categorizing each aspect of a privacy policy through visualization, more meaningful comparisons and analysis can be made than when reading a text-based privacy policy.



Fig 3.2.1 Examples of the Data Sharing and Legal Basis Aspects of the Just Eat Policy along with their respective colours.

3.2.2 Identification of Privacy Policy Issues

In order to successfully link multiple aspects together for meaningful analysis, issues must be recognised to act as an intermediary entity in the visualization. An issue is a sub-section of the privacy policy. While in this tool all use cases have the same thirteen aspects, privacy policies have no fixed amount of issues. The issues depend on the length and the detail involved in each document. Some examples of typical issues include:

- Information Provided
- Information Collected Automatically
- Information from Third Parties
- Processing Personal Information
- Data Retention
- Security
- Your Data Protection Rights
- Who Information is Shared with

While other “Issues” exist in some of the project use cases, the above are the main sub-sections which dominated the privacy policies that were selected for this project. In the PPVT a network diagram is displayed where issues and aspects are connected. An issue can have many aspects and an aspect can be associated with many issues, although issues are never linked to other issues and the same is true for aspects.

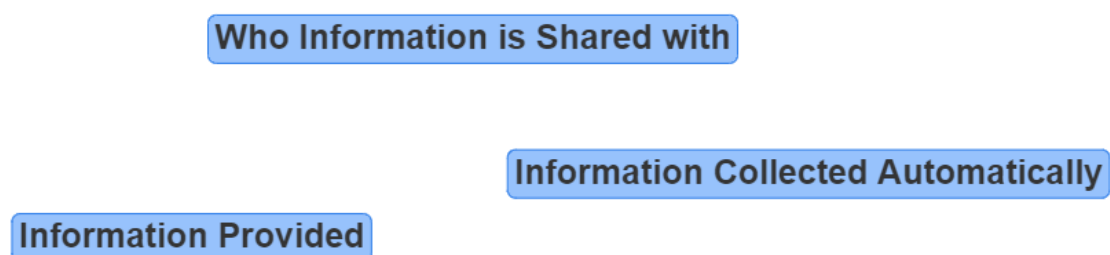


Fig 3.2.2 Examples of some of the Issues included in the Just Eat Privacy Policy

3.2.3 Layout and Structure

When designing the PPVT, it was vital to learn from the mistakes of other visualization tools. Users do not desire an overly complicated layout or too much visuals on the same page (Krempel, 2009), which is a mistake that many developers make when introducing a new visualization. The

most effective visualizations are quite simple in their design (Hanrahan et al., 2007), focusing more on conveying a certain message rather than appearing sophisticated and complicated. The layout and structure of the PPVT takes this into consideration. The white background, along with the breakdown of the visualization into multiple tabs ensures that users are not lost in a sea of unstructured data. The most efficient and effective way of separating the overall visualization was through designing each tab based on a specific aspect. As a result, there are fourteen screens, one for each of the aspects and an overall view. Following up on what has been mentioned before, issues and aspects are related. Each screen contains a unique aspect, but a certain issue may be present on multiple screens as a result of said issue involving multiple aspects of the categorization.

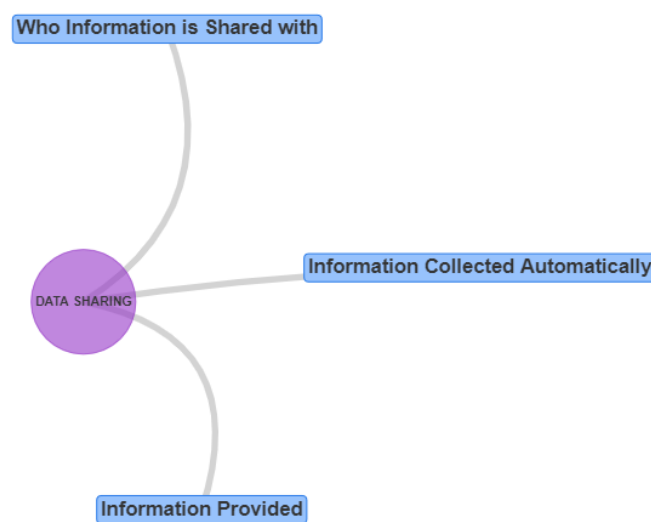


Fig 3.2.3 The relationship between Issues and the Data Sharing Aspect in the Data Sharing Tab of the Just Eat Policy.

The network diagram that is evident on each screen changes depending on the tab selected. If the data sharing screen is open, only this aspect and its related issues are drawn. The underlying software allows for both mouse and button navigation of the diagram, along with certain physics constants and parameters which increase interaction by allowing all nodes to move freely. Greater insight into these constants can be seen in the implementation section. Upon hovering over an issue, more information is displayed including any other aspects that are associated with the issue in question. While highly interactive, the visualization is also informative, as information and relative links complement the diagram on each screen.

3.3 Methodology

During the research phase of the PPVT it was vital to get an initial plan implemented which could be referred to and consulted with if the project ever hit any particular barriers. This project was introduced to provide users with a tool that informs users about their data and privacy, while also increasing the amount of interaction that the average individual normally has with the privacy policies of products and services. The system design diagram in the above section explains the methodology of this project while the below sub-sections demonstrate how each particular step was performed.

3.3.1 Annotating Privacy Policies

After identifying the types of aspects and issues that were to be collected in each privacy policy, the next task was to read through the documents and assign aspects and issues to the respective texts. In order to categorize the text into these aspects, annotation techniques were implemented. Annotation is a method of explaining or commenting on the detail of specific text. The most effective mode of annotation for these privacy policies was to highlight each section of text that related to a certain aspect in the colour that the aspect was associated with. Below is a table of the colours that were assigned to each aspect:

Aspect	Colour	RGBA
Data Type	Plum	(150, 50, 50, 0.4)
Data Category	Red	(255, 0, 0, 0.4)
Consent	Grey	(150, 155, 50, 0.4)
Rights	Pink	(255, 0, 255, 0.4)
Data Retention	Turquoise	(0, 255, 255, 0.6)
Location	Lilac	(150, 50, 125, 0.2)
Process	Blue-Grey	(0, 0, 250, 0.2)
Data Sharing	Purple	(150, 55, 200, 0.6)
Automated	Orange	(250, 175, 25, 0.6)
Data Source	Maroon	(128, 0, 0, 0.6)
Legal Basis	Sky Blue	(0, 0, 255, 0.6)
Processor	Yellow	(255, 255, 0, 0.8)
Third Party	Green	(150, 255, 0, 0.4)

Table 3.3 1 PPVT Aspects and Respective Colours.

By annotating each line of text and assigning these sections to certain aspects, the visualization became far easier to design. Although time consuming at first, annotating while reading the entire privacy policies ensured that no details were ignored or forgotten, while also making sure that the designer understood the policy and the type of document that the service provider was trying to create. Of the three privacy policies annotated, two were structured in such a way that issues could clearly be differentiated, leading to simple annotation. However, the issues still overlapped at times for one use case, which meant that the annotation process was still necessary.

For instance:

- **Registration information:** When you **create a Just Eat account**, sign-up or fill in forms on the Services, we **collect information about you** including your **name**, **address**, **email address** and the **password** you create.
- **Transaction information:** **We collect** information relating to your Orders, including **payment information** (e.g. your **credit card number**) using the secure services of our **processors**. Payment operations are outsourced to our **payment processors** and we do not store your credit card information in our systems. **We also collect** **delivery details** (e.g. your **physical address**) to fulfil each Order.
- **Information regarding your marketing preferences:** **We collect** information about your preferences to **receive marketing information** any time you subscribe or unsubscribe to our marketing.
- **Feedback:** When you **post messages and reviews** of the Services or you contact us, for example with a question, problem or comment, **we collect** information about your **name** and the **content of your query**. If you **contact our customer support teams** we will record and keep a **record of your conversation** for quality and training purposes and for the **resolution of your queries**.
- **Sensitive Information:** **We collect** information that you provide when you contact us (such as through our call centre or by using our online forms). This information may include sensitive personal information, such as **health-related information** (allergies or dietary requirements) or **information about your religion** (such as if you only eat halal food). We do not require this information, and we ask that **you share this information with the restaurant only**. However, there are some situations where you may nonetheless provide this type of information, for example if you make a complaint, and in those circumstances we will only be collecting it with your consent.

hide legend
data category
data type
process
automated
legal basis
data source
data retention
processor
third-party
data-sharing
consent
rights
location

Fig 3.3.1 A Section of the Just Eat Policy which has been annotated using the pre-determined aspects and their respective colours.

As can be seen in the above section of an annotated privacy policy, a legend has been included along the side of each document. These legends are fully interactive and are evident while scrolling through policies. This feature ensures maximum clarity regarding the aspects and their respective colours. The documents are readily available through a link in each of their associated interactive visualizations. The text-based annotated policies were included because, while clearly not suitable for most, certain individuals may prefer to understand and process information through text rather than visual displays. To ensure that all users gain valuable insight and informative knowledge on the subject, both the text-based annotated policies and the visualizations were always linked during this project.

3.3.2 Visualizing Privacy Policies

With the privacy policies annotated, the opportunity arose to visualize the policies. As mentioned above, a network diagram became the ideal visualization display considering the concept of representing all aspects and issues as being related in some shape or form. In the visualization,

each node was connected to at least one other node and the particular edges involved acknowledged the reference of node A in node B or vice versa. It was decided to have different shapes for aspects and issues, as can be seen from the above diagrams. Issues were drawn as boxes and aspects were drawn as circles. It was noted that users should be able to interact with the diagram with the assistance of both the keyboard and a mouse. As a result, buttons were introduced to the diagram which allow individuals to zoom and navigate through the aspects and issues.



Fig 3.3.2 A picture taken from the Data Retention tab. By hovering over the rectangular issue, other related aspects can be seen.

The concept of having the related aspects available upon request when hovering over a certain issue again reiterates the idea of connectivity and transparency between the different aspects and issues. It was considered that the aspects should be of different sizes depending on the amount of times they were referred to in the privacy policies. However, some aspects are mentioned far more often than others and as a result, certain nodes appeared visually inferior. Certain aspects were easily missed, and proportions were not as clear to the user. Having all aspects and issues the same size creates a somewhat symmetrical and aesthetic visual, which makes it far easier to analyse the information available.

3.3.3 Interacting with Privacy Policies

One of the main objectives for this project was to make privacy policies more interactive for the every-day user. It was noted from the beginning that individuals have no interest in reading privacy policies. Policies lack any sort of relativity or a method of interacting with the concepts and

messages conveyed (Zaeem & Barber, 2018). Each of the tabs in the PPVT offers a unique interactive opportunity with the respective privacy policies. While other aspects were considered such as a subsection of rights that focused more on GDPR and another screen that linked consent to processes, these additions held the potential to create uncertainty, with too much overlapping evident between screens and aspects. Below are the options that were decided upon, again making reference to the GDPRtEXT:

<i>Tab / Aspect</i>	<i>Interaction Opportunity</i>	<i>Reason / Motivation</i>
Data Type	Differentiate between data types (name, address, etc) that are provided voluntarily or collected automatically.	Users may find it useful to know which exact information they are providing to service providers.
Data Category	Differentiate between data categories (personal / financial information) that is provided voluntarily, collected automatically or provided by a third-party affiliate.	Being able to see what information is being provided by third-parties (Analytical Reports, Market Research, etc) can inform users of the objectives of the company in question.
Consent	Revoke consent from all available services.	The Consent tab gives users the opportunity to manage their consent while also showing which issues are affected by this, which is a feature that has never been available before.
Rights	<ul style="list-style-type: none"> • Unsubscribe from marketing communications. • Request that personal information is deleted. • Other features regarding the rights of the user. 	New standards such as GDPR have left people with the desire to manage their rights and other options through the same platform as the privacy policies in question.
Location	Gain an insight into which countries user information may be processed in.	Certain individuals may have an interest in the location tab due to countries having different policies and guidelines regarding privacy (Eg: Within / Outside EEA).
Data Retention	Object from the processing of personal information.	Some individuals might want to keep their account with a service open but not want to

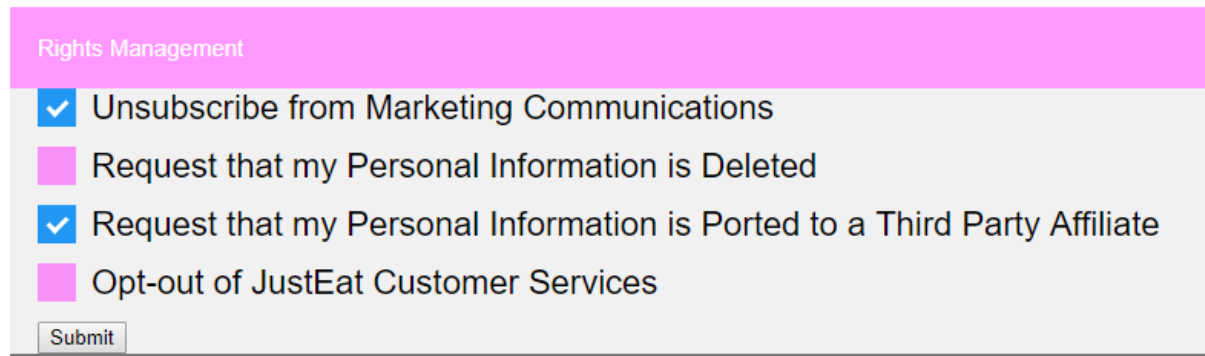
		have their personal information processed.
Process	Connect and make associations between the issues and processes of a particular privacy policy.	Being able to make this connection can be vital to truly understanding a privacy policy and the underlying motives of an organization.
Data Sharing	Differentiate between what information is shared between certain entities.	By having this opportunity, users can see first-hand where certain information is sent.
Automated	Gain an insight into how information is automatically collected from an individual's device.	By understanding this concept, users have more control over what information an organization can collect.
Data Source	Understand the underlying technology of cookies and how this affects the collection of data.	Again, this enables users to have more control over their respective information.
Legal Basis	Analyse the legal reasons for a number of important privacy policy processes (Eg: International Data Transfers, Information from Third Parties, etc).	Being able to understand why a company performs certain actions might increase trust levels between individuals and service providers.
Processor	Differentiate between processing that requires consent and processing that does not require consent.	Knowing exactly which issues require consent before the processing of information could also lead to improved trust levels.
Third-Party	Gain access to the privacy policies of third-party entities and affiliates.	Direct access to third party privacy policies gives users an increased understanding of how their data is collected, stored, managed and shared.

Table 3.3 2 PPVT Aspects and Interaction Possibilities

Just Eat Privacy Policy Visualization

[Show Standard Policy](#)

Rights



The screenshot shows a 'Rights Management' interface with a pink header. Below the header, there is a list of four items, each with a colored square icon (blue for checked, pink for unchecked) and a text label. The items are: 'Unsubscribe from Marketing Communications' (checked), 'Request that my Personal Information is Deleted' (unchecked), 'Request that my Personal Information is Ported to a Third Party Affiliate' (checked), and 'Opt-out of JustEat Customer Services' (unchecked). At the bottom left of the list is a 'Submit' button.

Rights Management	
<input checked="" type="checkbox"/>	Unsubscribe from Marketing Communications
<input type="checkbox"/>	Request that my Personal Information is Deleted
<input checked="" type="checkbox"/>	Request that my Personal Information is Ported to a Third Party Affiliate
<input type="checkbox"/>	Opt-out of JustEat Customer Services

Submit

Fig 3.3.3 An example of how the Rights Tab of the JustEat PPVT offers users with much more interaction and possibilities in comparison with standard policies or organization websites.

3.4 Technologies Used

As mentioned previously, some alternative technologies and techniques were considered when designing the PPVT and these designs possessed different technologies. While both effective in their own way, the finished product portrays the most informative and interactive tool, while also focusing on the user and the protection of their personal information. For this visualization tool a number of complementary technologies were used. While this tool does not possess the most sophisticated software in comparison with similar products such as Polisis or the Privacy Dashboard, the clarity and means in which the information is displayed has been prioritised to ensure the design of a product that individuals truly demand.

3.4.1 HTML & CSS

In regard to the design of the browser add-on, HTML and CSS were initially introduced. The use of HTML was first considered because of its ability to optimize the relationship between the visualization home page and the standard text-based policies. This simple markup language is still widely used and vital for any form of web development. HTML allows for seamless integration with other web design languages such as CSS and JavaScript libraries such as Vis.js. Similar to HTML, CSS is a styling language that can compliment any web development. The CSS involved in the PPVT ensures that features such as the navigation buttons, consent switches and service requests all portray a more professional level of design.

3.4.2 JavaScript & Vis.js

While HTML and CSS provide a foundation upon which a visualization tool can be built, there is a need for more technology to offer interaction functionality and a more dynamic web page. JavaScript is a high-level, interpreted programming language. It is characterized as dynamic and enables interactive web pages. When designing the PPVT, it was essential to include JavaScript in the chosen technologies in order to reiterate the goals and objectives of the project including making the tool interactive and easy navigation.

One of the most useful features of JavaScript is the many readily available libraries which allow for easier development and customization of JavaScript-based applications. Vis.js is a dynamic, browser-based visualization library. It was designed to handle large amounts of data and also to enable interaction with this data. The library consists of many components, including Timeline, Network, DataSet, Graph2d and Graph 3d. For the purpose of this project, the Network component was integrated and used to complement text-based privacy policies. Network is a visualization that is used to display data in the form of nodes and edges. It is easy to work with and runs smoothly on all modern browsers. The reason for choosing the Network visualization for this design was because the concept of using nodes and edges portrays the idea that while privacy policies may seem disjointed and unrelated in areas, surprising relationships and a more fluid policy are evident when these documents are visualized. Vis.js offers a great deal of interaction, which was also one of the primary features that similar tools lacked. In addition to being interactive and portraying connectivity, the library also offers some very useful physics and design constants which are fully customizable. These constants ensure that each visualization can be optimized for its respective policy. The library, while simple, was ideal for this type of project, bringing simplicity and sophistication together perfectly.

3.4.3 Alternative Technologies Considered

Implementing and designing visualizations is a very effective method of conveying a particular message or representing certain information in a new perspective. Their purpose is to offer a new means of accessibility and there are many alternative technologies that can be used when visualizing something like a privacy policy. The most integrative options come in the form of JavaScript libraries, as these allow for seamless incorporation with other useful technologies such as HTML, CSS and of course JavaScript. A shortlist of possible libraries were recognised early in order to select the most suitable candidates. Each mentioned had unique benefits, while also possessing limitations. D3.js was initially regarded as the most sophisticated library, offering a vast range of functionality and design features. However, adopting a nature of simplification and certain project requirements ensured that a less complex library may be more useful. Other libraries such as

MxGraph and Chart.js were consequently considered before it was agreed that Vis.js held the most potential as a visualization library.

After deciding on the optimum supporting library, it was important to investigate the relationship between data, the designer and the reader. The problem that this project aims to influence is the complexity of privacy policies and the confusion and misrepresentation of information that follow. By understanding how visualizations affect the involved entities, it became easier to reach a final design that fulfilled requirements. The Designer-Reader-Data Trinity is a concept that aims to clarify and solidify the relationship between entities involved, while also showing the effects that visualizations have on all three. The Trinity was created by Noah Iliinsky and Julie Steele in a book that seeks to demystify the design process associated with visualizations.

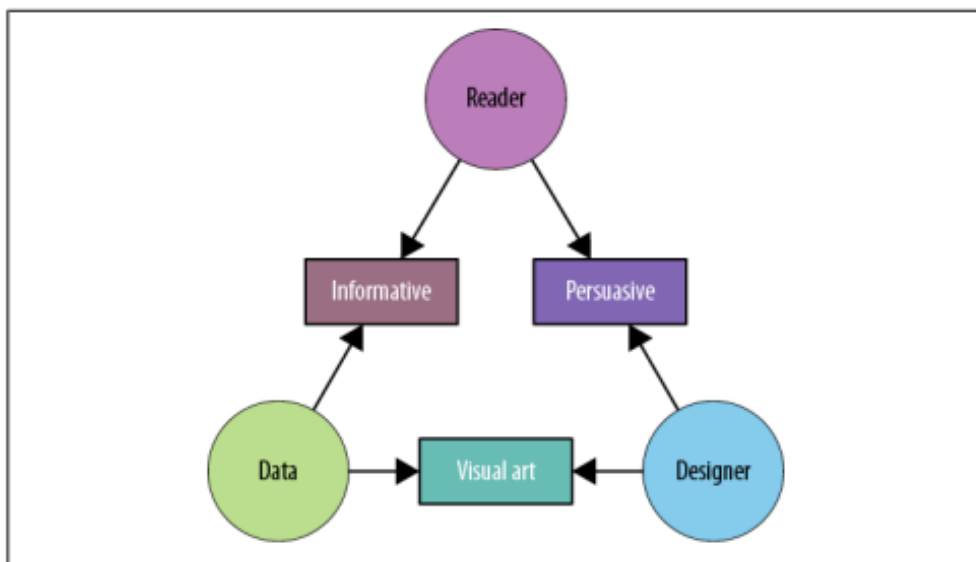


Fig 3.4.3 The Designer-Reader-Data Trinity (Iliinsky & Steele, 2011)

3.5 Use Cases and Visualization Examples

In order for the PPVT to be considered as a valuable and influential tool, the correct use cases had to be selected as examples. While any organization or product could be chosen, topical, complex and relevant cases provide the most insightful visualizations and offer the greatest improvement in comparison with standard text-based policies. In addition to this, it was important to note that the selected use cases would each alter the design and implementation of the PPVT. For example, certain privacy policies place more emphasis on consent, whereas others focus on data retention and how this affects users. As a result, the diagram can look very different, as different use cases will have unique proportions and distributions. Each visualization therefore has its own requirements and objectives. In a literary piece by Gottesdiener, the author refers to this fact and

also mentions that ignoring this concept can lead to a lack of quality along with customer dissatisfaction (Gottesdiener, 2003).

The use cases chosen for this project are BMW, Just Eat and Netflix. While each example was selected for a different reason, each case is valuable and together the three form very useful comparisons. As a world-renowned media service provider, Netflix is an organization that possesses a complicated privacy policy. This example is of great interest to the public, with the company having over 148 million subscribers currently worldwide (Statista, 2018). In addition, the idea of including an organization in the automobile industry was due to scandals in recent years such as the Volkswagen incident, whereby the German manufacturers installed devices which altered measurements to misinform individuals and testing agencies in the US (Bachmann et al., 2017). The public are in need of a product that improves relationships with car manufacturers and the PPVT may be considered as a possible solution. The final use case is JustEat. This organization was considered for the visualization tool because of the structure of its privacy policy. The simple layout and clear format of the Just Eat privacy contrasts with other policies and may prove to be more effective in the evaluation and user study stage of this project.

The privacy policies of BMW, JustEat and Netflix were visualized for the benefit of consumers and service providers alike. While these visualizations simplify and reduce content in comparison with text-based policies, this project is not intended to assist in the misrepresentation of information or to mislead users on their rights. During the initial phases of design this was a primary consideration, as a tool that can be used to manipulate customers through hidden small print and technicalities is ineffective and dangerous for all parties. Companies such as JustEat, Netflix and BMW are world renowned. Hence, these entities have quite complex and monumental operations in place. As a result, this project was never intended to become an integral part of every-day operations or have any effect on these organizations. Features of the PPVT such as the ability to revoke consent on services or the opportunity to unsubscribe from marketing communications are solely for demonstrational purposes regarding the possibilities that can come from such interactive tools and these features will not affect any users' relationships with these services.

4. Implementation and System Overview

4.1 Annotation to Visualization Decisions

The crucial step during the implementation of the PPVT was the transition between annotated text-based privacy policies and the visualization of these policies. This delicate process involved using the annotated documents as a primary foundation before adopting a mode of Motif Simplification to summarise and visualize the findings in a clear and uncomplicated manner. The possibility of over-simplifying said privacy policies was recognised as a potential danger and hence, a decision tree was implemented in order to decide whether certain pieces of information should be included or not.

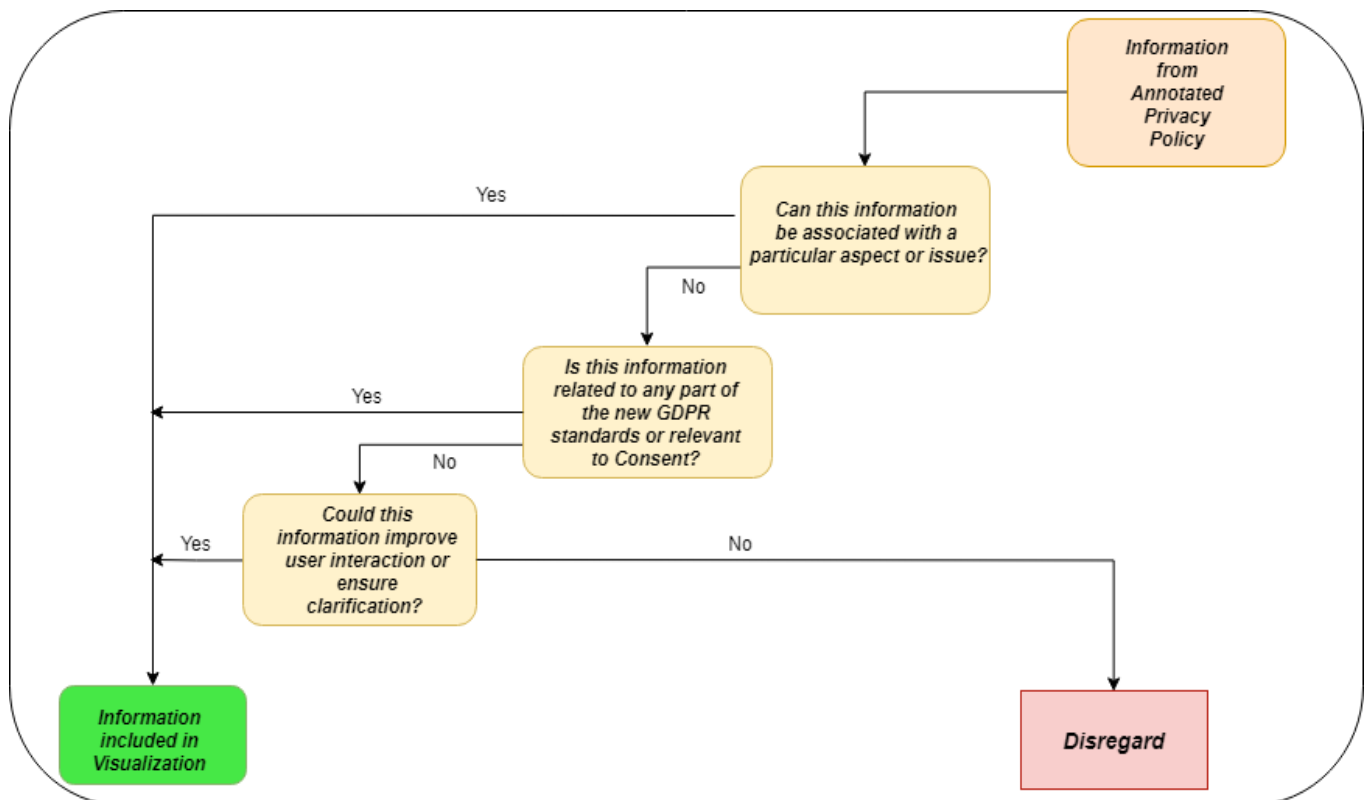


Fig 4.1 The Decision Tree which was used when choosing the correct information to include.

This decision tree became very useful, particularly when certain pieces of information appeared to exist between aspects or were considered to be in the grey area of policies. For example, the processes involved with the guidelines on fuel consumption and CO2 emissions in the BMW policy came into question. This information was not initially associated with any particular issue but is instead mentioned in a further piece of information at the bottom of the policy. While it is not related to any part of the new GDPR standards, it does ensure clarification regarding important information which has been questioned in the past and hence, was added to the

visualization. In comparison, the address of third-party entities such as Facebook or Instagram is not information that would be regarded as an improvement to user interaction or clarification. As a result, a link to their respective privacy policies is provided instead.

4.2 Network Physics Opportunities

One of the main benefits of using the Vis.js library during this project was the possibility to customize certain options and settings which are not available using other libraries. While elementary and simplistic in particular ways, Vis.js allows developers to alter some of the physics of nodes and edges which opens a window of opportunities in both design and implementation phases. Initial objectives such as increasing interaction opportunities and improving user understanding of privacy policies were greatly helped by these opportunities. As a dynamic and active type of visualization canvas, the Network component of Vis.js can seem difficult to manage at first. However, through the assistance of documentation it becomes evident that constants and variables can be altered and manipulated in order to create the optimal tool that fulfils both developer requirements and user needs.

4.2.1 BarnesHut Model

BarnesHut is a quadtree based gravity model. It is the recommended solver for non-hierarchical layouts such as is present in the Network component (Thieurmél, 2018). This model is based on an inverted gravity model and it offers a range of functionality and options including constants that affect how nodes move and how they interact with each other.

```
var data = {
  nodes: nodes,
  edges: edges
};
var options = {
  physics: {
    stabilization: true,
    barnesHut: {
      avoidOverlap: 0.8,
      springLength: 800,
      damping: 3
    },
    physics: false
  },
};
```

Fig 4.2.1 An example of some of the constants and customization possible with the BarnesHut model.

4.2.2 AvoidOverlap Constant

One issue faced during the implementation phase of the PPVT was the fact that at times the canvas became overcrowded with nodes. While Motif Simplification and the use of multiple screens assisted in the efforts to create a visually aesthetic tool, nodes and edges were still too compact, leading to a display that was difficult to comprehend. The introduction of the AvoidOverlap constant was invaluable to the display. By changing the value of this constant, nodes and edges were distributed to a greater extent across the canvas and the visualization as a whole became more understandable. As a result of initializing the AvoidOverlap constant to a number between 0 and 1, aspects and issues became more aware of their surroundings, which in tangent with gravitational numerical variables, ensured that these nodes were spread out and a more informative and clear visualization was designed.

4.2.3 SpringLength & Damping Constants

Almost similar to a domino effect, the AvoidOverlap constant led to further problems down the line. Due to the fact that nodes became too aware of other node existence and the constant was initially set at too high of a value, the aspects and issues in question were constantly in motion at one stage during the project. These nodes never settled, and the gravitational variables left the diagram moving in circular motion with objects avoiding each other. This error was fixed using two other constants; the SpringLength constant and the Damping constant. The SpringLength constant ensured that each edge was drawn at a minimal length, meaning that all aspects were of equal distance to their respective issues, while the damping constant limited unnecessary movement when an acceptable position was achieved.

4.2.4 Drawing Issues and Seed Allocation Constants

As a default setting, Vis.js Network diagrams are redrawn each time that changes are made or web pages are refreshed. While this may be useful for other projects, it introduced a lack of clarity in the PPVT as a result of aspects and issues being drawn in different locations on the diagram in multiple instances. This meant that the visualization would never look the same twice. The concept of simplification and the aesthetic JavaScript library tend to lose their effectiveness if familiarity and a clear visualization are not implemented. It became evident that it would be necessary to avoid redrawing the visualization after each reload. The RandomSeed constant was introduced which solved this problem immediately. This node allocation essentially keeps track of the way in which the Network diagrams are drawn and ensures that the design remains unchanged. By ensuring that diagrams have the same layout and structure every time, users are given the opportunity to become

familiar with specific privacy policies if they ever return to query a certain issue in the future. Products tend to experience increased customer value through repetition and familiarity (Trel, 2017), which is the objective for the PPVT.

5. Evaluation

5.1 User Evaluation

To measure the effectiveness of this tool, a user evaluation was conducted. A survey, along with the opportunity to interact with the visualization, was introduced to assess whether the PPVT assists users in understanding the control that they have over their own data. The study involved 50 participants and was implemented through an ethics procedure and application process. The questionnaire was available on Google Forms and this was used in coherence with the tool in the form of a browser widget to answer certain questions regarding usability, design and other aspects of the project.

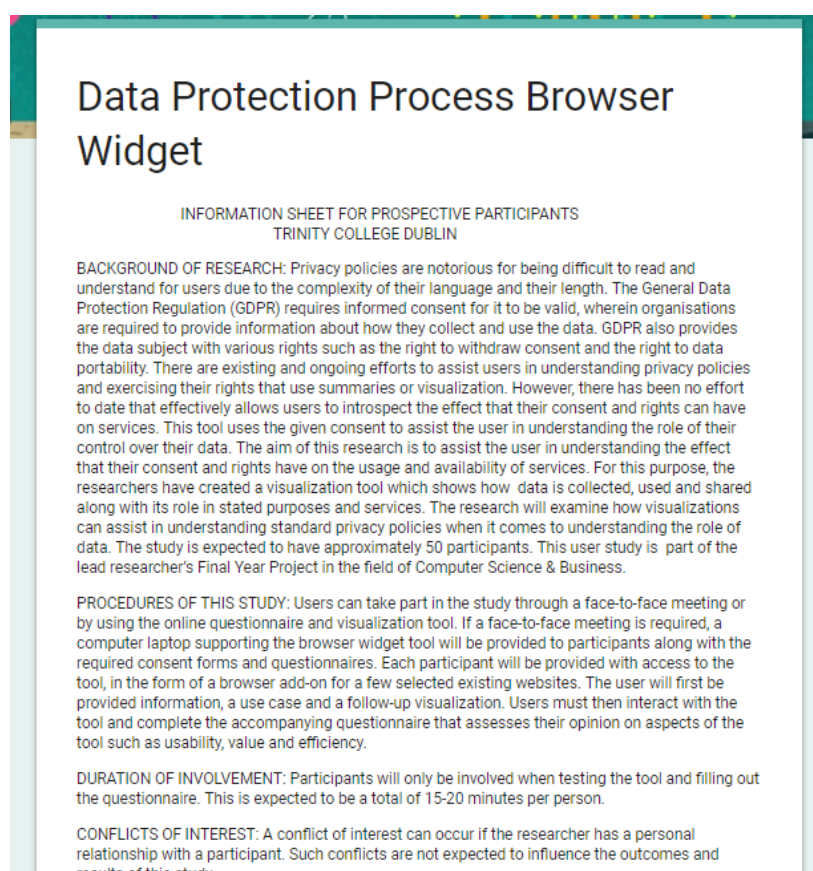


Fig 5.1 The Introduction Page of the User Evaluation

5.1.1 Consent

When including a user study in a project, certain standards and protocols must be followed. The TCD Research Ethics Committee ensure GDPR compliance and personal data protection by requiring consent forms to be included at the beginning of all project questionnaires and studies. A user can opt-out of or refuse to complete a survey at any stage during the process and can decline the opportunity to submit their answers at the end of the study.

RESEARCHERS CONTACT DETAILS:

INVESTIGATOR'S Name: James Cox

Date: 03/02/2019

Participant's Name:

Your answer

Date:

Your answer

Pre-Questionnaire Consent: *

☐ Yes, continue with the questionnaire

☐ No, stop and quit the questionnaire

BACK NEXT

Fig 5.1.1 1 The Consent Form Section of the Evaluation

If, at the beginning of a survey, a participant decides not to consent to the user study in question they are counted as a participant but their answers are recorded as invalid. During this user study, 2 participants decided against providing consent and so there were 48 valid responses.

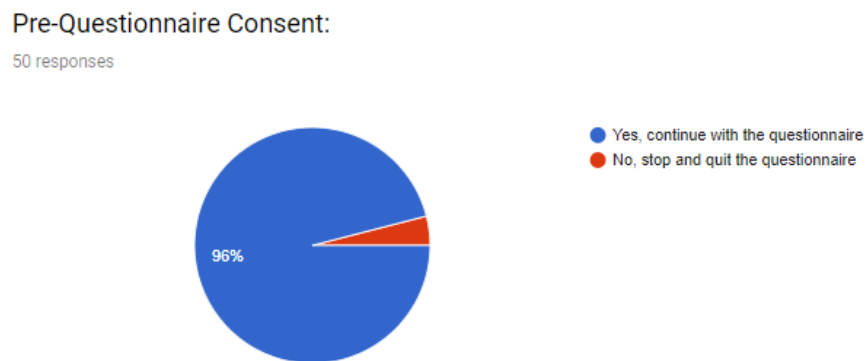


Fig 5.1.1 2 A Pie Chart representing Users and Consent for the Evaluation

5.1.2 Visualization Questions

During this evaluation, users were presented with two separate sets of questions. The first set of questions required individuals to interact with the PPVT to find certain pieces of information that were available within the visualization. Whether through interacting with aspects and issues or navigating through different screens, each answer was readily available and encouraged users to

improve their knowledge on the tool. Below is some of the questions in section 1 of the survey along with the responses that were received:

For how long does BMW retain most personal information?

- ☐ No longer than is necessary
- ☐ 90 days
- ☐ As long as is legally allowed
- ☐ As long as they want

Fig 5.1.2 1 Question 1 of the Visualization Survey

For how long does BMW retain most personal information?

48 responses

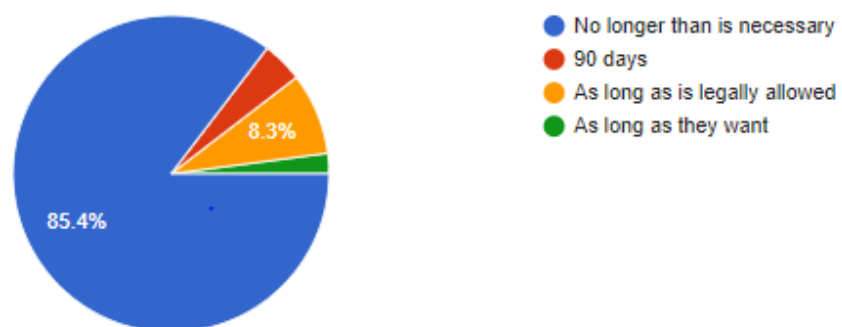


Fig 5.1.2 2 Response to Question 1

Which data type is automatically collected by Netflix?

- ☐ Payment Method
- ☐ Email Address
- ☐ Watch History
- ☐ Government Identification Number

Fig 5.1.2 3 Question 3 of the Visualization Survey

Which data type is automatically collected by Netflix?

48 responses

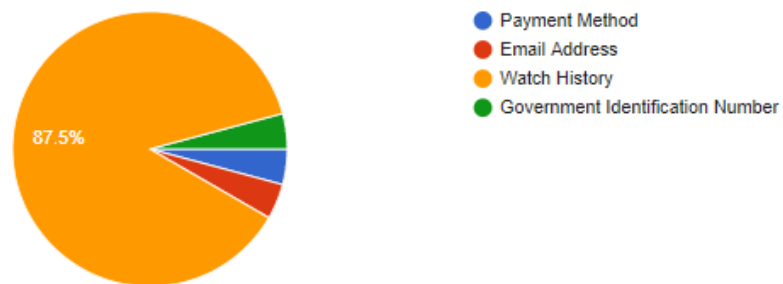


Fig 5.1.2 4 Response to Question 3

Just Eat: Identify some of the categories of personal information that are provided by you:

- ☐ Registration Information
- ☐ Delivery Details
- ☐ Activity Information
- ☐ Analytics Reports
- ☐ Religious Beliefs

Fig 5.1.2 5 Question 5 of the Visualization Survey

Just Eat: Identify some of the categories of personal information that are provided by you:

48 responses

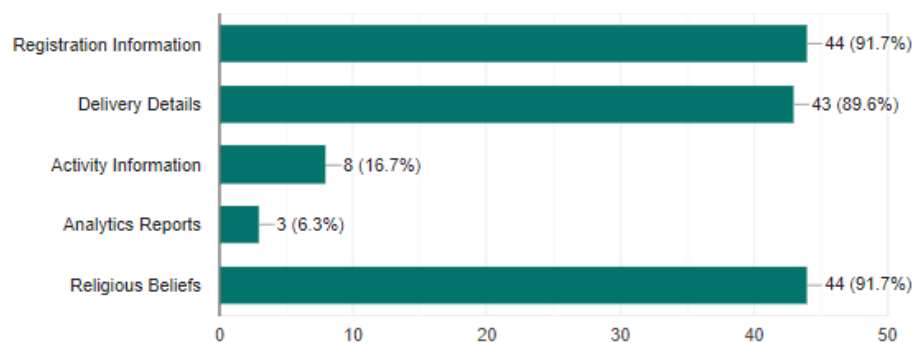


Fig 5.1.2 6 Response to Question 5

In total, 6 questions were asked during this stage of the evaluation. Each response contained a strong majority of correct answers with the lowest correct answer percentage being 79.2%, proving that users had little difficulty using the tool or dealing with the above questions.

Question	Correct Answer				
1	85.4%				
2	85.4%				
3	87.5%				
4*	91.7%	93.8%	89.6%		
5*	91.7%	89.6%	91.7%		
6	79.2%				

*Question involved checkboxes with three correct answers out of 5.

Table 5.1.2 1 Visualization Questions: Correct Answer Percentages

5.1.3 Systems Usability Scale

The second set of questions that were introduced to the users was based on the Systems Usability Scale (SUS). The SUS is a simple, Likert-scale questionnaire that provides an insight into an individual's opinion on the usability of a product or service. Unlike other SUS questionnaires, this evaluation contained 12 questions, while also offering individuals the opportunity to add further insight and to ask questions about the tool itself at the end of the survey. While users were not required to answer these questions, all 48 who answered the first set of questions also answered the SUS questions and hence offered some valuable insight. Below are some of the questions that were asked:

I think that I would like to use this tool frequently.

1

Strongly Disagree

2

☐

3

☐

4

☐

5

☐

Strongly Agree

Fig 5.1.3 1 Question 1 of the SUS

I needed to learn a lot of things before I could interact with this tool.

1 2 3 4 5

Strongly Disagree Strongly Agree

☐ ☐ ☐ ☐ ☐

Fig 5.1.3 2 Question 4 of the SUS

I became more informed about what happens to my personal data after using this tool.

1 2 3 4 5

Strongly Disagree Strongly Agree

☐ ☐ ☐ ☐ ☐

Fig 5.1.3 3 Question 10 of the SUS

The responses to the SUS questions such as the ones above are still useful towards the end of this project, particularly when focusing on future work and ways in which the PPVT can be optimized and developed. By including a user evaluation, researchers can gain first-hand opinions on the true value of a visualization tool which can be beneficial for all involved parties.

Question	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I think that I would like to use this tool frequently.	0%	4.1%	6.3%	41.7%	47.9%
I found the tool to be unnecessarily complex.	20.8%	66.7%	8.3%	4.2%	0%
I thought that the tool was easy to use.	0%	8.3%	2.1%	66.7%	22.9%
I think that I would need the support of a technical person to be able to use this tool.	37.5%	50%	6.25%	6.25%	0%
I would imagine that most people would learn to use this tool very quickly.	0%	2.1%	10.4%	68.8%	18.7%
I thought there was too much inconsistency in this tool.	27.1%	54.2%	12.5%	6.2%	0%
I needed to learn a lot of things before I could interact with this tool.	29.2%	58.3%	10.4%	2.1%	0%
I would make changes to the way in which the visualization is displayed.	16.7%	47.9%	29.2%	6.2%	0%

The visualization is generic and doesn't provide enough useful information.	18.8%	66.7%	10.4%	4.1%	0%
I became more informed about what happens to my personal data after using this tool.	0%	4.2%	4.2%	37.4%	54.2%
I found this tool to be useful when assessing user privacy.	0%	2.1%	12.5%	47.9%	37.5%
This tool provides a clearer insight into user data compared to a standard text-based privacy policy.	2.1%	4.2%	4.2%	62.4%	27.1%

Table 5.1.3 1 Results of the SUS

5.1.4 Ethical Considerations

Regarding the evaluation, the data and feedback that was collected during the user study and questionnaire will all be deleted upon completion of this project. This information was used for research purposes and anonymity has remained a priority during the course of this project. Both the visualization questions and the systems usability scale questions were designed to avoid any personal focus and therefore the project should encounter no ethical issues. This user study was introduced for the benefit of the researcher, but it was also implemented to increase user insight into the tool and show its future potential.

5.2 Privacy Evaluation

During the design and implementation phases of this project, creating a tool that was integrative and informative was the main priority. However, another highly relevant consideration was privacy. A privacy policy visualization tool that left users feeling vulnerable or insecure would effectively be worthless. While the tool must offer an insight into how information is collected, processed and shared, keeping this information solely for the purpose of user education and insight was vital. One particular method of measuring the privacy of this project was to implement a model based on the research of Kosa, El-Khatib and Marsh in which the group investigate the factors and states of privacy through different metrics and formulae. For the purpose of this project, a Privacy Indicator Metric (PIM) was designed. This metric will be used in order to measure the privacy and hence, in part the success of privacy policies, privacy tools and visualizations alike.

5.2.1 The States of Privacy

Kosa, El-Khatib and Marsh claim that there are essentially nine possible states of privacy in any computational environment (Kosa, El-Khatib and Marsh, 2011). The states are spread across the entire privacy spectrum and are a useful indication of whether a user can feel safe using a specific system or not.

#	State	Insight
1	Private	Total privacy. Existence is unknown.
2	Unidentified	Existence is unknown. Shadowy figure can be identified.
3	Anonymous	Limited information may be known. No link to specific identity.
4	Masked	Information is known but no links to identity are concealed.
5	De-identified	Information does not directly identify a person, but when linked with other information the person may become known.
6	Pseudonymous	Associated by an assumed (incorrect) name.
7	Confidential	Information revealed to an individual or organization acting in a certain role in a defined setting.
8	Identified	Information is capable of being distinguished and named.
9	Public	No private information. Complete openness.

Table 5.2.1 1 The States of Privacy (Kosa, El-Khatib and Marsh, 2011).

Creating a metric using the above states can give an indication of the privacy that users have when availing of certain privacy tools. A tool that is informative which also ensures a privacy state towards the lower end of the range is the ideal result for the PPVT. The PIM consists of three variables which take the form of human factors (H), data factors (D) and service factors (S). Each variable is of equal weighting with a maximum value of 3. The PIM is found by adding the three variables such that $P = H + D + S$, therefore giving a highest possible privacy value of 9.

5.2.2 Human Factors

The first variable that makes up the PIM is human factors. Kosa, El-Khatib and Marsh associate human factors with common privacy standards and properties of society. These so-called rules are factors that we sub-consciously associate with privacy and they can distinguish whether we consider certain products or services to be trustworthy or not. There are 15 factors that make up the human factors' variable.

Number	Factor	Consideration
H1	Object	Subject Matter?

H2	Appearance	Of Self, Others?
H3	Choice	Is Choice Possible?
H4	Control: Info	What Info is Disclosed?
H5	Control: Audience	Who is Present?
H6	Control: Access	Who may have Access?
H7	Discretion	Is Discretion Possible?
H8	Roles Established	Each Party has a Role?
H9	Status	Social Status or Invader?
H10	Common Bonds	Existing Relationship?
H11	Social Structure	Existing Social Relationship?
H12	Social Condition	What kind of Situation?
H13	Ritual Type	What are the Social Rules?
H14	Authority	Is there an Authority Figure?
H15	Public Expectation	Absence of Expectations?

Table 5.2.2 1 An adaptation of the Human Factors Set (Kosa, El-Khatib and Marsh, 2011)

As this variable is marked out of 3, the 15 factors make up 0.2 of the grading each. Policies, visualizations and other tools can be measured by grading each based on the above factors and marking the score out of a possible 3. Each tool is either marked as 0 or 0.2.

Number	Raschke's Privacy Dashboard	Pribo: Polisis	PPVT	Standard Policy (Netflix)
H1	0.2	0.2	0.2	0.2
H2	0.2	0	0.2	0
H3	0	0.2	0	0.2
H4	0	0.2	0	0.2
H5	0	0.2	0	0.2
H6	0.2	0.2	0	0.2
H7	0	0	0	0.2
H8	0.2	0.2	0.2	0.2
H9	0	0.2	0	0
H10	0.2	0.2	0	0.2
H11	0.2	0	0.2	0
H12	0.2	0.2	0	0.2
H13	0	0.2	0	0
H14	0	0	0	0
H15	0	0.2	0	0.2
Total	1.6	2.2	0.8	2.0

Table 5.2.2 2 Evaluation Table: Human Factors

5.2.3 Data Factors

The second variable of the PIM is the data factors variable. While data may not directly identify users, certain features and concepts such as provenance and transparency, along with new standards such as the GDPR ensure that the below data factors are influential towards the privacy of users.

Number	Data Type	Examples
D1	Biographic	Name, Age, Birth Date, Race, Ethnicity, Religion, Marital Status
D2	Demographic	Mailing Address, Telephone Number, Location History
D3	Health	Diagnosis, Care Plans, Genetic Information, Blood Type, Test Results
D4	Financial	Credit Card Numbers, Account Information, Credit History
D5	Provenance	Service Provider, User
D6	Third Party	Affiliates, Subsidiaries
D7	Identifiers	Driver's License, Health Care Numbers, Medical Record Numbers
D8	Behavioural	Preferences and Choices, Current Physical Location

Table 5.2.3 1 An adaptation of the Data Factors Set (Kosa, El-Khatib and Marsh, 2011)

Similar to the human factors set, the data factors variable can be a maximum of 3. As there are 8 factors, each tool is graded as either 0 or 0.375 for each specific factor.

Number	Raschke's Privacy Dashboard	Pribo: Polisis	PPVT	Standard Policy (Netflix)
D1	0.375	0	0.375	0.375
D2	0.375	0	0.375	0.375
D3	0	0	0	0
D4	0	0	0	0
D5	0.375	0.375	0	0.375
D6	0	0.375	0.375	0.375
D7	0	0	0	0
D8	0.375	0	0.375	0
Total	1.5	0.75	1.5	1.5

Table 5.2.3 2 Evaluation Table: Data Factors

5.2.4 Service Factors

The final variable associated with this metric is the service factors value. Computers are generally accepted to be effective tools for information management. Machines can now be used to read information without human intervention. However, certain tools still remain more private than others regarding certain factors.

Number	Factor	Sources of Identifiable Information
S1	Network	User, Machine
S2	Hosting	User, Machine

S3	Registration	User, Machine
S4	Messaging	User, Machine
S5	Backup	User, Machine, Metadata (both)
S6	Software	User, Machine, Metadata (both)
S7	Interaction	User, Machine, Metadata (both)
S8	Websites / Portals	User, Machine, Metadata (both)

Table 5.2.4 1 An adaptation of the Service Factors Set (Kosa, El-Khatib and Marsh, 2011)

The service factors variable is analysed similarly to the data factor values. As there are 8 factors and a maximum value of 3, each factor carries a weight of 0.375.

Number	Raschke's Privacy Dashboard	Pribot: Polisis	PPVT	Standard Policy (Netflix)
S1	0.375	0	0.375	0
S2	0	0	0	0
S3	0.375	0	0.375	0.375
S4	0	0.375	0	0.375
S5	0.375	0	0.375	0
S6	0.375	0	0.375	0
S7	0	0.375	0	0.375
S8	0	0.375	0	0.375
Total	1.5	1.125	1.5	1.5

Table 5.2.4 2 Evaluation Table: Service Factors

Privacy Tool	Raschke's Privacy Dashboard	Pribot: Polisis	PPVT	Standard Policy (Netflix)
Human Factor (H)	1.6	2.2	0.8	2
Data Factor (D)	1.5	0.75	1.5	1.5
Service Factor (S)	1.5	1.125	1.5	1.5
PIM (H+D+S)	4.6	4.075	3.8	5
Privacy State*	De-identified	Masked	Masked	De-identified

* Rounded to Nearest Whole Number

Table 5.2.4 3 Evaluation Table: Privacy State

5.3 Evaluation Summary & Limitations

The creation of the PIM has been included to complement the findings of the above user valuation and to offer a valuable contrast between the PPVT and other products and services. As can be seen from the above tables, the PPVT offers a very similar, but lower PIM score and privacy state than other visualization tools when the three factors are considered. As a result, the tool is classified under the masked privacy state, resulting in a tool that could be considered to have a higher regard for privacy compared to other services. Although the PIM is a measure that was only designed for this project and solely used for research purposes, it allows for a somewhat meaningful comparison between tools and interestingly conveys that standards text-based policies hold a similar level of privacy to visualization tools. However, the standard policies still lack interaction, and so the necessity for the alternative products remains.

Regarding the limitations of the evaluation, the PIM still lacks certain alternative considerations and there remains the possibility of adding other factors to increase accuracy. Many high-profile individuals such as Mark Zuckerberg, Eric Schmidt and others have recently claimed that privacy is dead and that there is no benefits or even point in attempting to measure it (Dorraj, 2014). The idea of building a metric to measure privacy has been regarded by the above as a neutral process which should neither prove or disprove the opinion of others about privacy and relevant products.

6. Conclusion

6.1 Objective Assessment

As mentioned above in the introductory section, there were some clear goals and objectives established at the beginning of this project. While on a limited time schedule and acknowledging the need to place equal emphasis on alternative final year modules, the above objectives appear to have been achieved to the best of the researcher's ability. The objectives were as follows:

1. Establish a visual design that improves the user's understanding of privacy policies while still displaying all necessary information.
 - As is evident from section 3, a design was introduced which aims to provide all necessary details while simplifying the respective policies. The results of the user study reinforce the fulfilment of this objective, with user's understanding of the policies in question clearly improving after using the tool.
2. Increase emphasis on consent and rights, particularly regarding GDPR changes, compared to other products and services.
 - One particular aim for the project was to make the PPVT as interactive as possible. When it was decided that the visualization would have separate screens for each aspect, it became necessary to include some sort of interaction on each screen. Multiple screens offered the opportunity to revoke consent from particular service providers. The 'Rights' tab, for example, allowed users to investigate their GDPR rights while also altering the other functionalities of each service.
3. Conduct a user evaluation in order to measure the effectiveness of the tool and also to gain an insight into user opinion of the visualization.
 - A user study was successfully implemented which gave a valuable insight into the public's opinion on the PPVT. In addition, a research evaluation was conducted which compared the tool to similar products, thus offering an alternative method of measuring the effectiveness of the product.

6.2 Project Issues

As mentioned, the development of the PPVT and the project as a whole was satisfactory and relatively efficient. However, certain problems arose. One particular issue in the design phase was the fact there was no possible way to overlap the Network canvas with the HTML development. As a result, some of the screens display visualizations of different sizes to others. This had to be

accounted for because of the need for interactive features which limited the amount of screen space on certain pages. However, the visualizations still convey the required message while also increasing the possible interaction with the user.

In addition, there were some limitations to the amount that could be achieved with this project as a result of time constraints and the other modules that were chosen this year. The achievement of project objectives is due to maintaining organization between modules and introducing planners that allocated time evenly between different assignments. The project itself had to be broken down into sub-sections which ensured that there were no processes or stages that received more time than others. A Gantt chart, along with the web application Trello, was used to keep track of necessary processes and meetings with both the supervisor and the assistant supervisor.

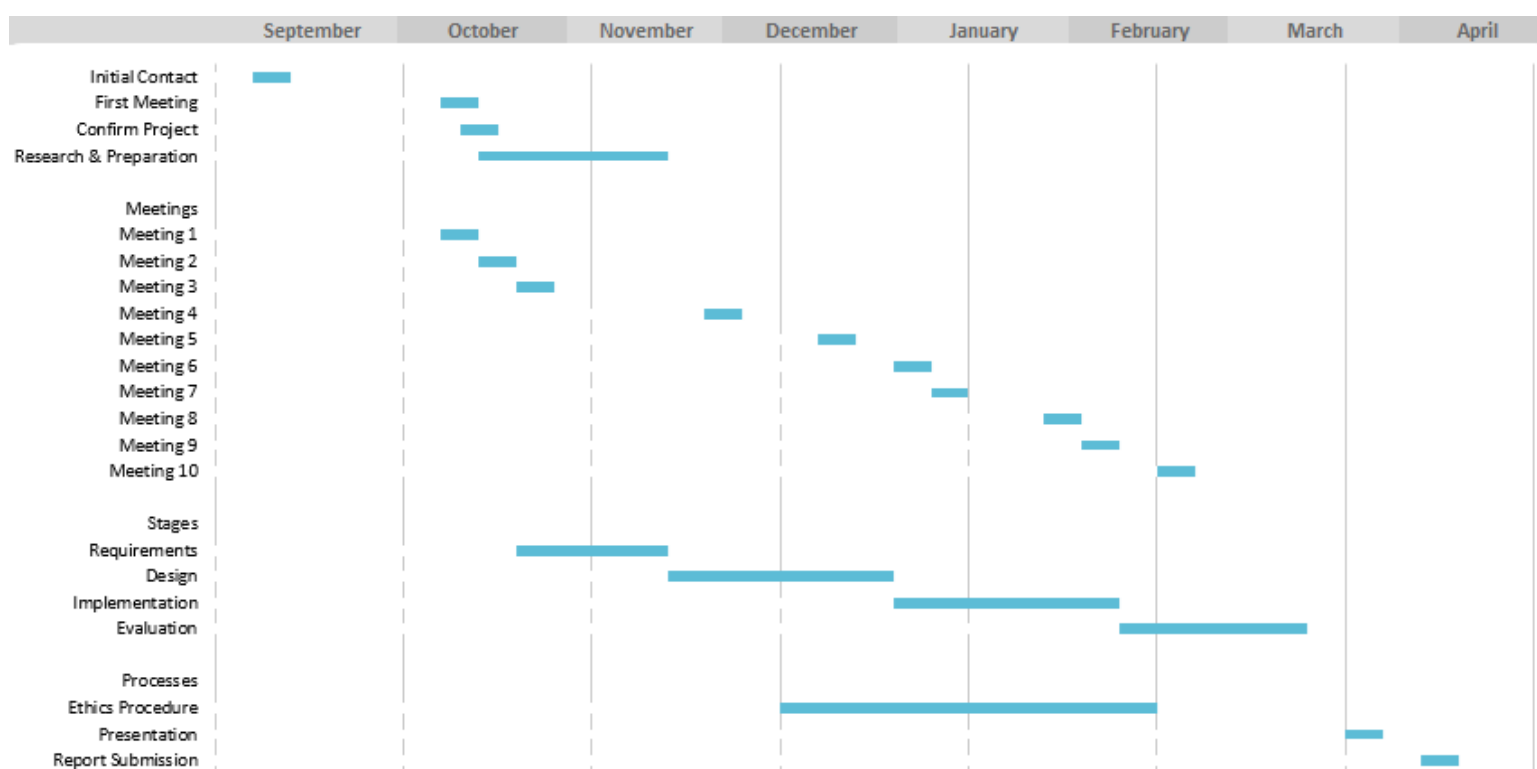


Fig 6.2 Gantt Chart introduced during this project.

6.3 Future Work

Ideally, future work would include further development of the PPVT. As a result of time constraints and the fact that this was a final year project, there were numerous technologies and features that the researcher did not have the possibility to include. The number of use cases would

initially be increased as a temporary fix on the amount of privacy policies currently visualized. With a more long-term outlook, the researcher would have liked to investigate the possibility of introducing alternative technologies such as machine-learning, which holds the potential to be integrated in an attempt to design a visualization tool that could perform both the annotation and visualization stages automatically through sophisticated software and algorithms. Regarding the business perspective for the project, if the PPVT ever became sophisticated enough to be integrated into real world systems and products, the opportunity to reach out to clients such as JustEat or Netflix could be something worth considering.

Similar to the work of both the supervisor and the assistant supervisor, the lead researcher found the topic of data provenance, with a particular focus on GDPR, to be particularly interesting, and if the opportunity ever arose in the future to study in this line of work, it would also be a considerable opportunity.

Bibliography

- 1) Bachmann, R., Ehrlich, G. & Ruzic, D. (2017). *Firms and Collective Reputation: the Volkswagen Emissions Scandal as a Case Study*. [online]. Available at: https://www3.nd.edu/~rbachman/BER_current.pdf [Accessed 03 Mar. 2019].
- 2) Burgess, A. (1986) But Do Blondes Prefer Gentlemen? *Homage to QWERT YUIOP and other writings*. [online] Available at: https://kisslibrary.net/book/12C993CC55A04EB054AC?utm_source=dude&utm_medium=scalesup.wfnen.org&utm_campaign=fnom&x=536947 [Accessed 12 Mar. 2019].
- 3) Council of the European Union, (2015). *Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*. [online]. Available at: <http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf> [Accessed 02 Apr. 2019].
- 4) Dhotre, P. S. (2017). Systematic Analysis and Visualization of Privacy Policies of Online Services. [online]. Available at: http://vbn.aau.dk/files/268003536/PHD_Prashant_Shantarman_Dhotre_E_pdf.pdf. [Accessed 31 Mar. 2019].
- 5) Dorraji, S.E. & Barcys, M. (2014). *Privacy in Digital Age: Dead or Alive?* [online]. Available at: <https://www.mruni.eu/upload/iblock/b97/ST-14-4-2-05.pdf> [Accessed 10 Mar. 2019].
- 6) Dunne, C. & Shneiderman, B. (2012). *Motif Simplification: Improving Network Visualization Readability with Fan, Connector, and Clique Glyphs*. [online]. Available at: <http://www.cs.umd.edu/hcil/trs/2012-29/2012-29.pdf> [Accessed 12 Mar. 2019].
- 7) European Parliament Policy Department, (2015). *Surveillance and Censorship: The impact of technologies on human rights*. [online]. Available at: [http://www.europarl.europa.eu/RegData/etudes/STUD/2015/549034/EXPO_STU\(2015\)549034_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2015/549034/EXPO_STU(2015)549034_EN.pdf) [Accessed 04 Apr. 2019].
- 8) Gottesdiener, E. (2003). *Use Cases: Best Practices*. [online]. Available at: <http://www.eg.bucknell.edu/~cs475/F04-S05/useCases.pdf> [Accessed 08 Mar. 2019].
- 9) Hanrahan, P., Stolte, C. & Mackinlay, J. (2007). *Visual Analysis for Everyone*. [online]. Available at: http://cdnlarge.tableausoftware.com/sites/default/files/whitepapers/visual_analysis_for-everyone.pdf [Accessed 15 Mar. 2019].
- 10) Harkous, H., Fawaz, K., Lebrete, K., Schaub, F., Shin, K. & Aberer, K. (2017). *Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning*. [online]. Available at: https://pribot.org/files/Polisis_USenix_Security_Paper.pdf [Accessed 18 Mar. 2019].

- 11) Iliinsky, N. & Steele, J. (2011). *Designing Data Visualizations*. [online]. Available at: http://courses.ischool.utexas.edu/unmil/files/Designing_Data_Visualizations.pdf [Accessed 21 Mar. 2019].
- 12) Kis, V. (2010). *Learning for Jobs OECD Reviews of Vocational Education and Training*. [online]. Available at: <https://www.oecd.org/ireland/44592419.pdf> [Accessed 16 Mar. 2019].
- 13) Krempel, L. (2009). *Network Visualization*. [online]. Available at: <http://www.mpifg.de/~lk/netvis/onlinepdf/Visualization53aJB2.pdf> [Accessed 22 Mar. 2019].
- 14) Kurtz, C., Semmann, M. & Bohmann, T. (2018). *Privacy by Design to Comply with GDPR: A Review on Third-Party Data Processors*. [online]. Available at: https://www.researchgate.net/publication/325415927_Privacy_by_Design_to_Comply_with_GDPR_A_Review_on_Third-Party_Data_Processors [Accessed 24 Mar. 2019].
- 15) Lamba, S., Jacob, J. & Rawat, A. (2014). *Impact of Teaching Time on Attention and Concentration*. [online]. Available at: <http://www.iosrjournals.org/iosr-jnhs/papers/vol3-issue4/Version-1/A03410104.pdf> [Accessed 25 Mar. 2019].
- 16) Lukács, A. (2016). *WHAT IS PRIVACY? THE HISTORY AND DEFINITION OF PRIVACY*. [online]. Available at: <http://publicatio.bibl.u-szeged.hu/10794/7/3188699.pdf> [Accessed 28 Mar. 2019].
- 17) Marotta-Wurgler, F. (2015). *Understanding Privacy Policies: Content, Self-Regulation, and Market Forces*. [online]. Available at: https://www.law.uchicago.edu/files/file/marotta-wurgler_understanding_privacy_policies.pdf [Accessed 02 Mar. 2019].
- 18) McDonald, A. & Cranor, L. (2008). *The Cost of Reading Privacy Policies*. [online]. Available at: <http://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf>. [Accessed 06 Mar. 2019].
- 19) NIHRC, (2008). *Inspiring Practice: Resources, tools and activities for human rights education*. [online]. Available at: http://www.nihrc.org/uploads/publications/Inspiring_Practices.pdf [Accessed 08 Mar. 2019].
- 20) Pelteret, M. & Ophoff, J. (2016). *A Review of Information Privacy and Its Importance to Consumers and Organizations*. [online]. Available at: <http://www.inform.nu/Articles/Vol19/ISJv19p277-301Pelteret2584.pdf> [Accessed 12 Mar. 2019].
- 21) Raschke, P., Kupper, A., Drozd, O. & Kirrane, S. (2017). *Designing a GDPR-compliant and Usable Privacy Dashboard*. [online]. Available at: <https://www.specialprivacy.eu/images/documents/IFIP-2017-Raschke.pdf> [Accessed 19 Mar. 2019].
- 22) Silvianita, Khamidi, M. F. & Kurlan, V.J. (2013). *Decision Making for Safety Assessment of Mobile Mooring System*. [online]. Available at: <http://personal.its.ac.id/files/pub/5211->

silvianita-oe-01-Jurnal%20Teknologi,%20ISSN%200127-9696,%20Vol%2066%20No%203%20.pdf [Accessed 16 Mar. 2019].

- 23) Solin, B., Sara, B. & Vanda, V. (2017). *Customer Uncertainty: The relationship between psychic distance and consumer behaviour towards purchasing from foreign online retailers*. [online]. Available at: <http://www.diva-portal.se/smash/get/diva2:1120612/FULLTEXT02.pdf> [Accessed 18 Mar. 2019].
- 24) Statista, (2018). *Number of Netflix streaming subscribers worldwide from 3rd quarter 2011 to 4th quarter 2018*. [online]. Available at: <https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/>. [Accessed 20 Mar. 2019].
- 25) Stephenson, J. (2015). *Abuse and Misuse of Personal Information*. [online]. Available at: <https://www.alec.org/app/uploads/2015/11/Abuse-and-Misuse-of-Personal-Info-Final-03202013.pdf> [Accessed 21 Mar. 2019].
- 26) Syncopation Software, (2018). *DPL™ 9 Fault Tree User Guide*. [online]. Available at: https://www.syncopation.com/downloads/DPL_9_FT_User_Guide.pdf [Accessed 24 Mar. 2019].
- 27) Thieurmél, B. (2019). *Package 'visNetwork'*. [online]. Available at: <https://cran.r-project.org/web/packages/visNetwork/visNetwork.pdf> [Accessed 28 Mar. 2019].
- 28) Trel, M. (2017). *The Effect of Product Familiarity on Consumers' Attention to Online Advertisements*. [online]. Available at: <https://pdfs.semanticscholar.org/81d4/953ccbbede268f243919c8aa1d8159d9d9df.pdf> [Accessed 25 Mar. 2019].
- 29) Verizon, (2018). *2018 Data Breach Investigations Report*. [online]. Available at: http://www.documentwereld.nl/files/2018/Verizon-DBIR_2018-Main_report.pdf [Accessed 31 Mar. 2019].
- 30) Waldman, E. (2018). *PRIVACY, NOTICE, AND DESIGN*. [online]. Available at: https://law.stanford.edu/wp-content/uploads/2018/03/Waldman_Final_031418.pdf [Accessed 9 Mar. 2019].
- 31) Zaeem, R. N. & Barber, S. (2018). *A Study of Web Privacy Policies Across Industries*. [online]. Available at: <https://identity.utexas.edu/assets/uploads/publications/A-Study-of-Web-Privacy-Policies-Across-Industries.pdf> [Accessed 12 Mar. 2019].

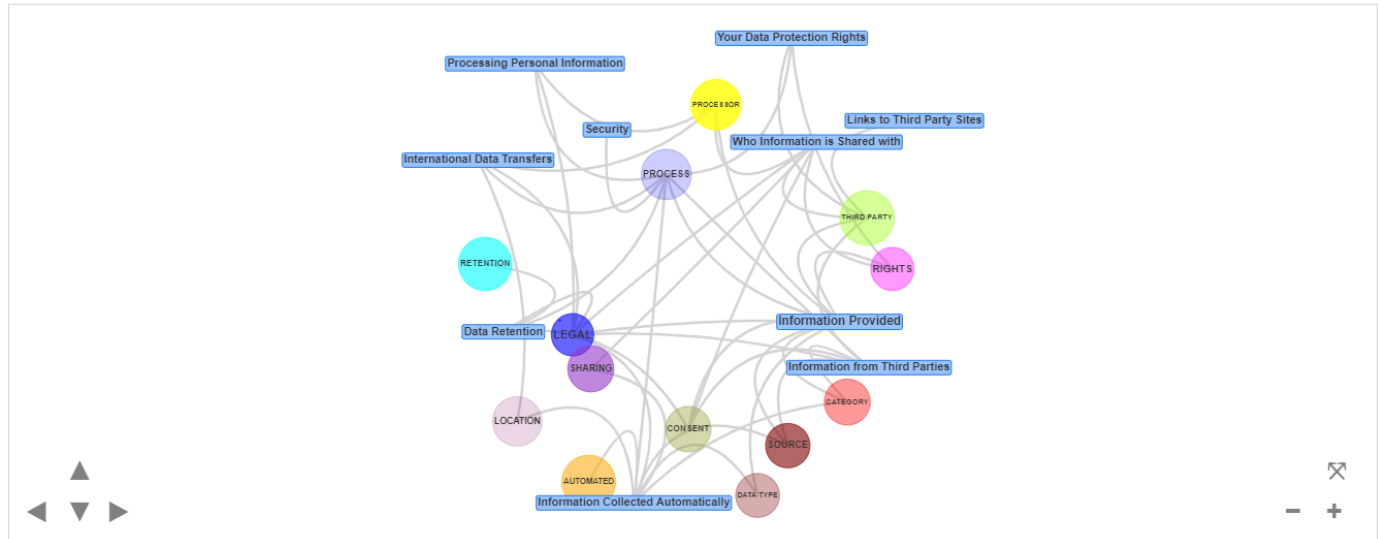
Appendices

Appendix A: Overall View of the PPVT (Netflix)

Netflix Privacy Policy Visualization

[Show Standard Policy.](#)

Overall	Consent	Rights	Retention	Process	DataSource	Processor	Automated	Location	DataType	DataSharing	ThirdParty	LegalBasis	DataCategory
---------	-------------------------	------------------------	---------------------------	-------------------------	----------------------------	---------------------------	---------------------------	--------------------------	--------------------------	-----------------------------	----------------------------	----------------------------	------------------------------



Appendix B: Consent Tab of PPVT (BMW)

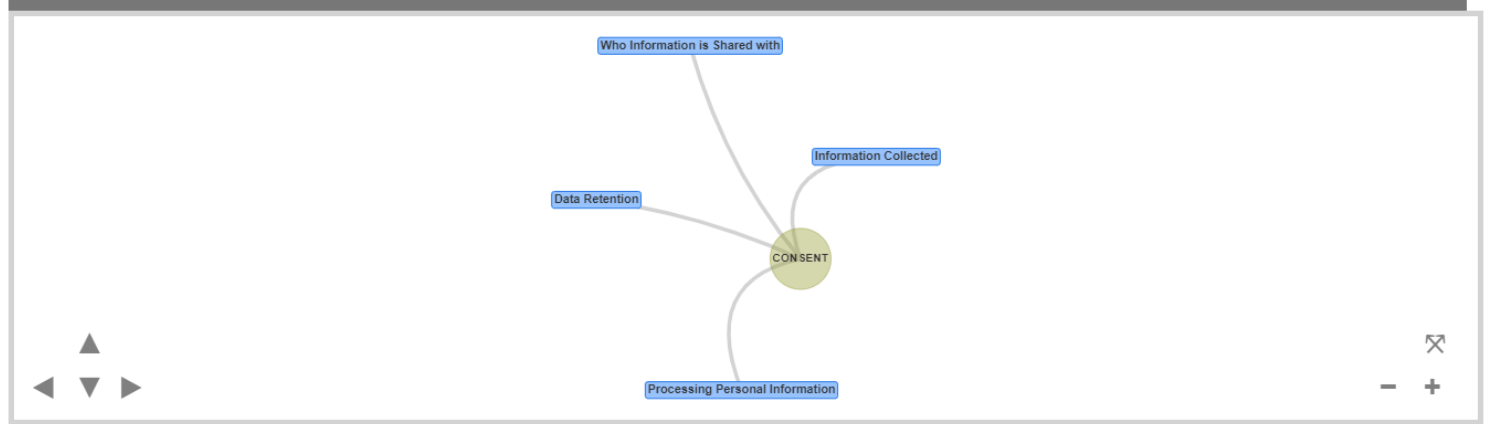
Overall	Consent	Rights	Retention	Process	DataSource	Processor	Automated	Location	DataType	DataSharing	ThirdParty	LegalBasis	DataCategory
-------------------------	-------------------------	------------------------	---------------------------	-------------------------	----------------------------	---------------------------	---------------------------	--------------------------	--------------------------	-----------------------------	----------------------------	----------------------------	------------------------------

Consent

[Revoke Consent on all BMW Services](#)



Issues affected by Consent



Appendix C: Rights Tab of PPVT (JustEat)

Just Eat Privacy Policy Visualization

[Show Standard Policy](#)

[Overall](#) [Consent](#) [Rights](#) [Retention](#) [Process](#) [DataSource](#) [Processor](#) [Automated](#) [Location](#) [DataType](#) [DataSharing](#) [ThirdParty](#) [LegalBasis](#) [DataCategory](#)

Rights

Rights Management

Unsubscribe from Marketing Communications

Request that my Personal Information is Deleted

Request that my Personal Information is Ported to a Third Party Affiliate

Opt-out of Just Eat Customer Services

Submit

Your GDPR Rights

Your Data Protection Rights

RIGHTS

Appendix D: Evaluation Submission Page

Data Protection Process Browser
Widget

This Questionnaire is now complete

Your help with this questionnaire has been greatly appreciated, thank you.

BACK

SUBMIT

Never submit passwords through Google Forms.