PROJECT PROPOSAL

Master of Software Engineering Project Part A – 2021

Tianqing Gao, Chirag Garg

The University of Adelaide | School of Computer Science

Table of Contents

Introduction **2**

Project Scope **3**

Background and Related Work **5**

Data Privacy Background: *5*

Related Work *5*

Research Method: **8**

Proposed Solution: **9**

Project Plan: **10**

Introduction

Leaking of Privacy has been a huge problem recently. A global survey found that 88% of people are concerned that their privacy may be leaked and 80% of people expect the government to regulate the companies' use of technologies and fine those companies that do not use the personal data properly [5,11]. To protect privacy and our personal information that the companies collect, Privacy Impact Assessment (PIA) is a process that companies should conduct before the development starts. Privacy Impact Assessments (PIAs) provide a structured risk analysis of a system's privacy impacts on a set of individuals and offer recommendations for managing the risks. It is a process that needs to be held before the development starts. Different organizations need it to not only ensure compliance but also to identify appropriate privacy practices and controls. This allows the organisation to have a transparent and accountable set of privacy practices and that would help them gain public trust and confidence. However, Privacy Impact Assessments can be complex and time-consuming for those companies who do not have any experience in this field before. Flaherty [1] has already criticized the existing PIA guidance documents, arguing that they should be simple and easy to understand, but the truth is that most of the documents are still complex and not "business friendly" to those companies that need to conduct the PIA. Since the PIA process is mostly a volunteering effort, too complicated steps may make organisations give up on conducting the process. This absence of any privacy risk analysis may increase the probability of privacy violations in the future. The Australian government's Privacy Act requires agencies to conduct a PIA for all high-risk projects that could impact the privacy of individuals. The agencies can have a preliminary threshold assessment that determines a project potential privacy impacts and whether it is a high-risk privacy project. Then they could decide to have a PIA under 'Office of the Australian Information Commissioner' (OAIC) rules.

New technologies are collecting users' data to provide better service but on the other hand the data flow of the private data has not been transparent, increasing the public's privacy concerns. Companies also refrain from doing a PIA because the cost of time and money on it may be too high. However, according to Flaherty's practical experience [1], the cost of conducting a PIA could be less than 1% of the whole developing process. It can also prevent the organizations from paying fines due to non-compliance that are caused by misuse of users' personal data. Without the PIA, the developing team might also use the data in an illegal way, causing major losses to the company in the future and the app could not be released and would waste the company a lot of money. Also, the public would lose trust in the company, damaging its reputation. With our tool, the companies can streamline the PIA process and avoid the privacy risks that the application would face as well as any legal issues related to privacy in the future. We in our project are planning to make a PIA tool that is compliant with the Australian regulations. It could be a web-based or stand-alone tool. The tool would have the ability to extend compliance with other countries regulations as well.

Not many tools that assist users to do the PIA have been developed. We only found a tool that is developed by the French team. Compared to their tool, we would introduce a tool for Australians, and we would make the user interface more user friendly and allow the users to choose based on their decision. To make the PIA process easier for the companies, we are going to make a tool that can help them do the PIA quicker and in a more straightforward way.

Project Scope

Project Scope Statement	
Project Name	Privacy Impact Assessment Tool for Australian Organisations

Scope Description	IN SCOPE:
	• A web-based system to help users carry out Privacy Impact Assessments.
	• Create a user-friendly interface for the PIA's privacy threat modelling.
	  ○ Allow the user to register a new PIA.
	  ○ Allow the users to select all the relevant Privacy Targets for their assessment.
	  ○ Allow the users to enumerate Privacy Threats and classify them in terms of likelihood and impact (e.g., low, medium high).
	  ○ Allow the users to enumerate Privacy Controls as countermeasures to the enumerated threats.
	  ○ Allow the visualization of a PIA Report of the entire assessment.
	• Guide the users to do the PIA in a way that is easy to understand.
	• A guide video to show how to use our tool.
	• A list of regulations that allow users to choose from.
	
	OUT OF SCOPE:
	• Providing other language choices for the users.

Project Deliverables	• A web interface for PIA tools to be made available as an open-source artifact. • A complete report and documentation of the developed tool. • A README file to use the PIA tool.
Work Description	To build a web-based tool that helps the users to do the PIA process easier.
Constraints	No customer service or other language options to choose other than English.
Assumptions	The user will be familiar with Australian privacy regulations.

Project Requirements:

- Users:

  The potential users of our product can be the Data Privacy Officer (DPO) of the company or organization, it can also be developers who want to demonstrate their compliance.

- User stories associated with the users:
  - As a DPO, I want to be able to sign up so that next time I can login with the user information.
  - As a DPO, I want to be able to change my user profile so that I can add description to the project being processed.
  - As a DPO, I want the product to have a share function so that I can share the current PIA that I am conducting with my colleagues.

- As a DPO, I want to be able to track the history of the PIA I have processed in my user profile so that i can know what i have done before.
- As a DPO, I want to be able to continue what I haven't finished last time so that I don't need to refill the fields I have filled before.
- As a DPO, I want to be able to change the regulation I want to follow and not refill the same fields so that I don't need to redo the same work after changing.
- As a DPO, I want to be able to change the previous PIA I have conducted so that I can have a new version of the PIA process without filling some same information.
- As a DPO, I want to have the option to choose from a variety of regulations so I can choose the ones I want to follow based on our requirements.
- As a DPO, I want to have a guide video so that I can know how to get started.
- As a developer, I want to be able to see what is left to be filled so that I won't forget to fill in if I skipped one part.
- As a DPO, I want the download PIA report clear and business-friendly so that I can use the PIA report in practice.
- As a DPO, I want to be able to receive the PIA report via email so that I can have a backup of the report in my mail.
- As a DPO, I want to be able to reset my password so that when I forget my password, I can still have a way to assess my account.
- As a DPO, I want the website to have authentication when login so that it will be more secure and other people will find it hard to steal the account.
- As a DPO, I want to be able to choose multiple regulations to follow so that our application can meet different regions' data protection policy.
- As a DPO, I want to be able to write a description of the Project in my history of conducting PIA so that I can know which report is related to which project based on the description.
- As a DPO, I want to be able to receive a reminder email if I haven't finished the PIA process, I started so that I won't forget to finish the undone report when I am busy.

- Use Case:

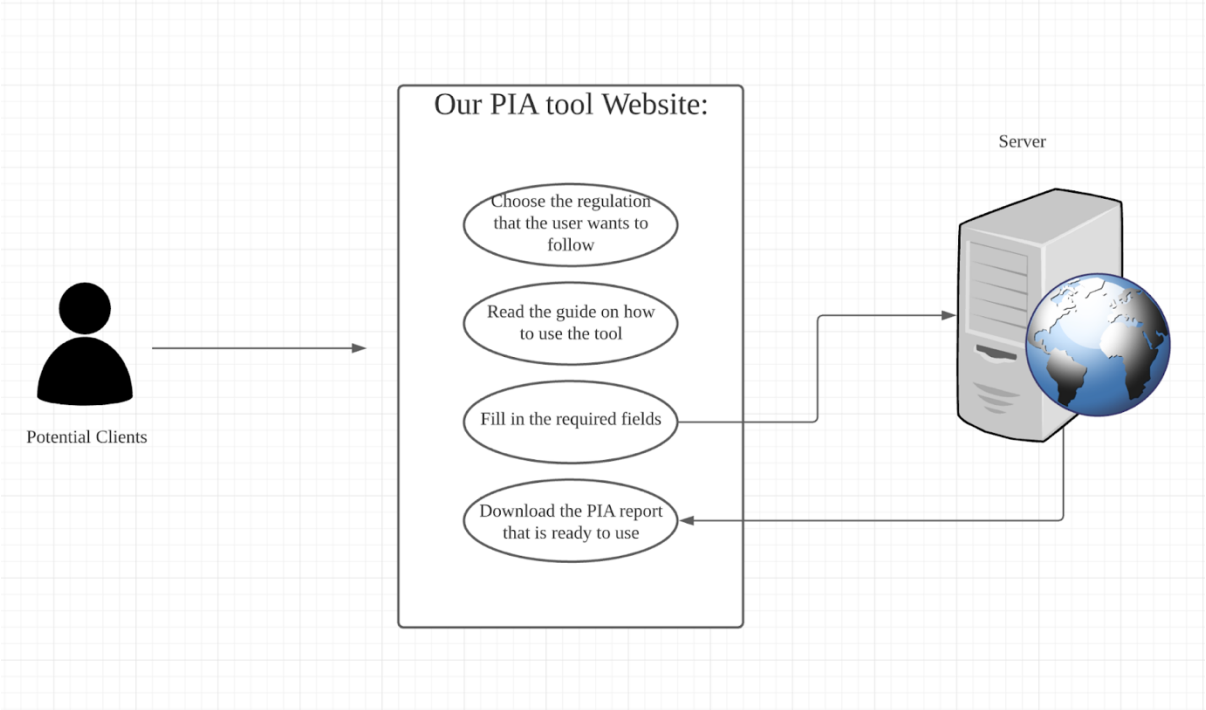

Figure 2.1

Background and Related Work

PIA Background:

PIA is a concept started from the idea of "impact assessments "in the 1970s [5]. It is aimed to conduct an assessment to find out the potential privacy problems and define controls in the early stage of the development [4]. More and more countries and regions release their own version of the PIA guidance documents, while in Australia, for most states, there is only the Data Protection Law Compliance Assessment, but we cannot just simply refer them to PIA. Until 2009, Victoria released their own guidance documents to help the individuals and organizations to conduct their own PIA. The documents' quality is quite high and has been accepted by many researchers who are conducting research in this area. The only problem Clarke [3] mentioned is that the structure and description of the document makes it more like a project not a process.

When preparing the PIA, companies should identify the main security problems related to private information. Then following the PIA guidance documents to finish the PIA process. The core part of the PIA process is the risk assessment. Risk assessments have a process of justification and mitigation. PIA should follow the same idea of that [5]. Since the project may get updated and have more functions added to the product, multiple PIAs need to be conducted [1]. European "RFID PIA template" proposes a PIA process in four steps [5]: (1) describe the system landscape, (2) identify the potential threats, (3) use appropriate controls to mitigate the threats, (4) document the analysis result and the threats found into a PIA report. Multiple guidelines have been released, and the most famous one among them is the GDPR one, which is a good one that we can make use of when conducting our PIA methodology.

Data Privacy Background:

Data Privacy

PRIVACY AS CONFIDENTIALITY

A common conceptualisation of privacy, in a technical re-interpretation of the user right to be left alone privacy term, is to avoid making personal information available to any individual, especially the general public. The goal of privacy technology, according to this concept, is to allow the use of services while minimising the amount of information that is revealed. Both clear data shared with the provider and information made indirectly accessible in the Metadata associated with these exchanges are referred to as information in this context [12].

PRIVACY AS TRANSPARENCY

Transparency systems, as opposed to applications that restrict data disclosure or use of exposed data, examine users' online behaviours in order to either provide input on the consequences of their actions or perform audits to ensure that there has been no infringement of privacy. Transparency-based privacy, like control-oriented systems, cannot avoid privacy abuses on its own. In reality, reviews or audits take place after users have already provided the provider their data. As a result, providers are once again entrusted with ensuring that the collected data is not stored or exchanged in ways that the users have not approved [12].

PRIVACY AS CONTROL

A broader meaning of privacy, which is sometimes cited in legislation, expands the idea of privacy beyond the concealment of personal information to the right to monitor what happens to that information after it is released. The idea behind the change away from technologies that restrict transparency and towards technologies that allow users to monitor how their data is used is that in many situations, disclosing data is inevitable or viewed as advantageous to the data topic. As a consequence, it is advisable to consider the use of technologies that address two major concerns: I enable users to communicate how they expect data disclosed to the service provider to be used, thus preventing unnecessary processing of these data; and ii) enable organisations to identify and implement policies that prevent the misuse of information, as defined by users [12].

PRIVACY ENGINEERING (Privacy by Design)

Unlike conventional risk analysis, which protects a company's properties, the GDPR protects data subjects' rights and freedoms, including the right to data privacy and the right to full control and information of their personal data. The GDPR defines Data Protection Impact Assessment (DPIA). DPIA is a risk-based approach to improving and demonstrating compliance with these criteria. There is a suggestion of a three-step approach for performing the DPIA and include a supporting tool. There is also focus on risk analysis as a component of this approach. The provided tool aids controllers in facilitating the rights and freedoms of data subjects. The work is distinguished from others by the assistance provided by the tool [6].

The General Data Protection Regulation (GDPR) encourages organisations to use Data Protection Impact Assessments (DPIAs) to incorporate privacy into their activities and practises from the start. To date, however, there has been little guidance on how Protection & Privacy Requirements Engineering processes map to the DPIA's needed activities and how these activities can be assisted by tools. There is a tool-supported method for conducting DPIAs using existing Requirements Engineering approaches and the CAIRIS framework to address this problem. There is use of a real-world case study example to demonstrate this technique, which was used to evoke privacy risks for a prototype medical application to support chemotherapy care [7].

It is a challenging task to build ubiquitous computing systems that are consistent with data protection regulations. According to the European General Data Protection Regulation, device developers must use a privacy-by-design approach and conduct privacy impact assessments during the implementation life cycle. The proposal is a software-assisted process architecture that makes privacy implications in ubiquitous computing systems easier to analyse. Students and ubicomp experts have tested this programme [8].

There is a look at the criteria for conducting Privacy Impact Assessments (PIAs) in a cloud computing environment and how a PIA support tool could be designed. In cloud computing, privacy is a critical concern, as real or perceived privacy flaws can affect legal enforcement, data protection, and user trust. A privacy impact assessment (PIA) is a structured method for determining the potential future consequences of a particular action or initiative on an individual's privacy. It focuses on understanding the method, initiative, or scheme, detecting and minimising negative privacy impacts, and educating policymakers who must determine whether or not to continue with the project in its current form. As a proactive business procedure, a PIA is distinct from reactive procedures such as privacy problem review, privacy audits, and privacy law enforcement testing, which are applied to existing programmes to ensure that they continue to comply with internal and external laws [9].

Prior to beginning a potentially risky processing operation, the General Data Protection Regulation (GDPR) mandates that organisations perform Data Protection Impact Assessments (DPIA or PIA). The PIA tool is a standalone and "server" version of a free and

open-source software tool. It assists businesses in performing PIAs by directing them through the process step by step, enabling them to demonstrate GDPR compliance.[10]

Related Work

	Paper 1	Paper 2	Paper 3	Paper 4	Paper 5
Architecture	/	/	Software-assisted	Cloud Environment	/
Technology	RiskML Tool	CAIRIS Framework	PATH Framework	JAVA,HTML,JavaScript	/
PIA Methodology	Three-step DPIA methodology	Tool-supported DPIA	DSRM	/	Step by Step approach
Relevant Laws	GDPR	GDPR	GDPR	Not Specified	GDPR
Open source	Yes	Yes	Yes	Yes	Yes

In [6], there is no specified architecture mentioned, the technology used is 'RiskML' tool. There is also a three step DPIA methodology used. Here it follows the European guidelines namely GDPR. And also, the tool here is open source.

In [7], there is also no specified architecture mentioned, the technology used is 'Computer Aided Integration of Requirements and Information Security' (CAIRIS) Framework. There is also a tool supported DPIA methodology used. Here it follows the General Data Protection Regulation (GDPR) as well. And this is open source as well.

In [8], there is a software-assisted architecture, the technology used is Privacy Aware Transmission Highway (PATH) process framework. There use of Design Science Research Methodology (DSRM). Here it follows GDPR as well and is open source.

In [9], there is a cloud environment architecture, the technology used are JAVA, HTML and JavaScript. There is no PIA methodology specified here. And it is also does not follow any relevant laws. It is open source as well.

In [10], there is no mention of a architecture and neither of the technology used. But it follows a step-by-step PIA methodology. It also follows the European GDPR and is open source.

Research Method:

For our project, multiple research methods can be applied to it. They can be divided into two different parts; some will be applied before the developing process and some of them will be applied after the development. The methods applied before the developing start is to identify the users' requirements and have a better understanding of the scope and topic of the whole project. And the ones to be applied after the project is done is to check the usability and how our tool works in the practice. Mainly, we plan to use document analysis, opinion survey, systematic review. and controlled experiment as our research method.

By applying the document analysis, we will go through the different data protection policies among the different regions and countries. For example, the most famous one from GDPR and the PIA guidance documents from OAIC. By analysing these documents and policies, we can find the difference and that will help us to design the components of our PIA Tool. The only limitation of this method is that the amount of the documents may not be that much. But with the limited number of documents, we could still learn a lot and have a better understanding of the scope of our project.

Opinion survey is another method that could be applied in this project. By doing a survey among the small number of possible end users, we can have a deeper understanding of the user requirements and what are their expectations of the Tool. That will also contribute to the development of the tool. Knowing the user's needs can help us know what we need to embed in our product. The only limitation of that is the number of users who take the survey won't be too high and that could cause bias on the survey results.

Another research method we could apply is the systematic review. By critically reviewing the existing literatures on privacy, risk management, impact analysis, we will know exactly how the PIA process should be done and can come up with an improved solution that will meet the requirements of the Tool.

After the Tool is developed, controlled experiments could be conducted. We can have two groups of people, one group familiar with the PIA process and another isn't familiar with it to see the difference of the time they spent on the PIA report. Another experiment can be asking the groups to do the PIA report by filling the document and by using our Tool. By comparing the time spent, we can find out if our Tool helps them to do the PIA report. With this method, we can evaluate our work and find out where we can improve in this product.

Proposed Solution:

In our project, we are expecting to make a web-based application. The application would have the functions that help the users to conduct their PIA process. Users will be able to sign in and access the web site with the user information. The application would be able to let the users to choose which regulation or multiple regulations users want to follow. Based on the choices, the fields that are required to be filled would be different. After the user follows our guidance and fills all the blanks then a PIA report that is ready for the user to download would be available for the users. they can either choose to get the report from email or simply download the report from the website.
Below is the flow chart of the PIA process that we will follow:

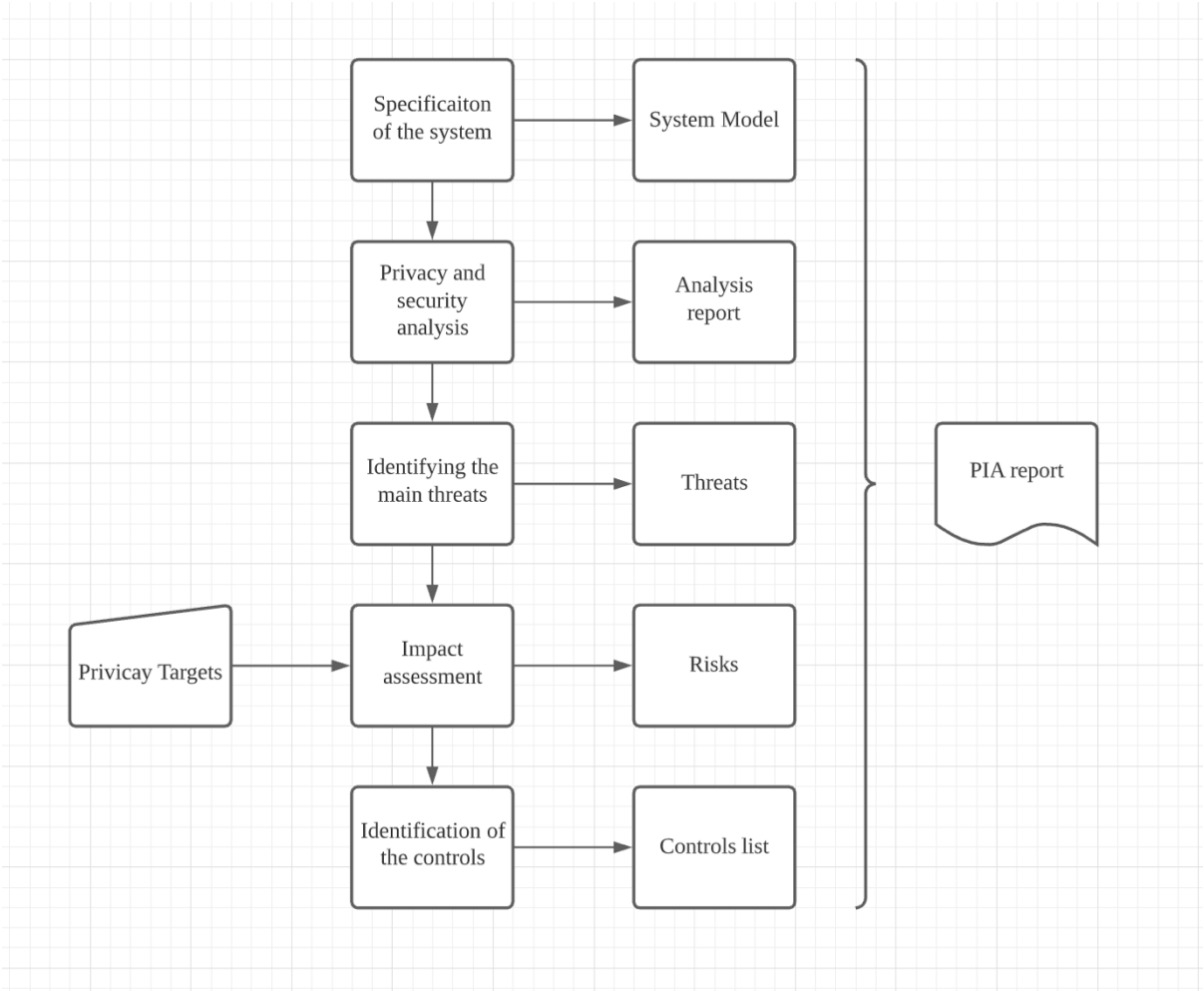

Figure 4.1

To develop this application, we will use JavaScript, Node.js, HTML and CSS. This is because we are building a web-based application. HTML, CSS and JavaScript will be used for the front-end design and Node.js would be used for back-end development. The reason why we choose to develop a web-based application is that the web interface is more user friendly and no need to download any application. All the work can be done online. It's easy for the potential clients to access and more likely for those organizations to find our application. Another technique we need is a database. We will choose MySQL as our database tool, because both of us have some experience using it. The reason why we are using a database is that we want to collect the data we get from the user and use the information in the database to produce

the report. Also, we can get track on the previous PIA reports we did for the clients. For the deploy tools, we would be using webhosting or the AWS cloud to do the work. Because basically our tool is a web-based application, with the tools we choose we can collect the data online to do the production of the PIA report. For development and operations, we would choose GitHub to do that. We choose this because both of us are familiar with GitHub and we can do project management stuff on GitHub to get track on what we have done and what we are going to do and what we are working on right now.
The following is the software architecture for our project:

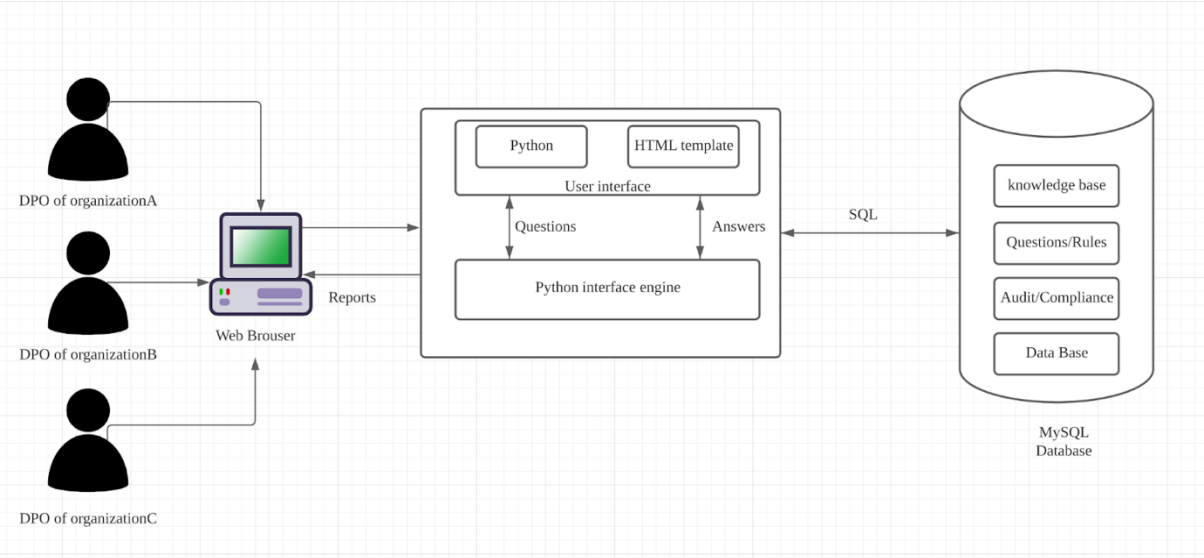

Figure 4.2

Project Plan:

Our research will be divided into three main parts: researching, designing and developing. For the research part we have read some of the Privacy Impact Assessment papers. The literature review is divided into two parts. One of us is responsible for the papers that are talking about the methodology foundations for PIAs and the other one focuses on the concrete tools that have already existed. We would compare the existing PIA Tools and compile a list of relevant features for our proposal. After we have done the basic research of the PIA related papers and tools, we are expecting to summarize what we got and come to a conclusion on the detailed development plan. Also, we will briefly design the main pages of the product using mock-ups. At the developing stage, we would make it an open-source project on GitHub, anyone who is interested in the project can do their own contribution for the project. This semester we are planning to build a demo for our product and perfect our product in the second semester. The detailed plan for this semester is as follows. The first two sprints we are focusing on the literature review part and the basic design part and the next two sprints we will be focusing on the development of the PIA Tool. The detailed schedule is as follows:

Sprint number	Tasks to be performed	Who will be responsible for the tasks?
Sprint1(~3.28)	1. Research on Privacy Impact Assessment methodology and foundations 2. Research on the existing tools on Privacy Impact Assessment and compare the difference between them. 3. Compare existing PIA Tools and compile a list of relevant features for our proposal.	Task 1,3 assigned to Tianqing. Task2 assigned to Chirag.
Sprint2(~4.11)	1. Define high-level architecture of the entire platform. 2. Define detailed use case diagrams. 3. Briefly design the main pages of the product using mock-ups. 4. Perform the initial data modelling for the database, identifying main entities and data items. 5. Setup the development environment. 6. Build main pages of the PIA Tool.	Task 1,2,3 assigned to Chirag. Task 4,5,6 assigned to Tianqing.
Sprint3:(~5.9)	1. Design the login functions and connect the user information to the database. 2. Design the user profiles for the users. 3. Design the login page. 4. Develop the sign in function. 5. Build the sign-up page of the PIA Tool. 6. Build the login page of the PIA Tool.	Task 1,2,3 assigned to Chirag. Task 4,5,6 assigned to Tianqing.
Sprint4:(~5.23)	1. Define the fields that different regulations need to follow. 2. Develop the user history function. 3. Develop the authentication function. 4. Develop the password reset function. 5. Design the documentation page of the PIA Tools.	Task 1,2 assigned to Tianqing. Task 3,4,5 assigned to Chirag.

References:

1. Flaherty, David. "Privacy impact assessments: an essential tool for data protection." *Privacy Law & Policy Reporter* 5 (2000): 85.
2. Wright, David. "Making privacy impact assessment more effective." *The Information Society* 29.5 (2013): 307-315.
3. Clarke, Roger. "An evaluation of privacy impact assessment guidance documents." *International Data Privacy Law* 1.2 (2011): 111-120.
4. Ahmadian, Amir Shayan, et al. "Supporting privacy impact assessment by model-based privacy analysis." *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. 2018.
5. Oetzel, Marie Caroline, and Sarah Spiekermann. "A systematic methodology for privacy impact assessments: a design science approach." *European Journal of Information Systems* 23.2 (2014): 126-150.
6. Dashti, Salimeh, and Silvio Ranise. "Tool-Assisted Risk Analysis for Data Protection Impact Assessment." *IFIP International Summer School on Privacy and Identity Management*. Springer, Cham, 2019.
7. Coles, Joshua, Shamal Faily, and Duncan Ki-Aries. "Tool-supporting data protection impact assessments with CAIRIS." *2018 IEEE 5th International Workshop on Evolving Security & Privacy Requirements Engineering (ESPRE)*. IEEE, 2018.
8. Fernández, Alfredo Pérez, and Guttorm Sindre. "Software Assisted Privacy Impact Assessment in Interactive Ubiquitous Computing Systems." *Conference on e-Business, e-Services and e-Society*. Springer, Cham, 2019.
9. Tancock, David, Siani Pearson, and Andrew Charlesworth. "A privacy impact assessment tool for cloud computing." *Privacy and security for Cloud computing*. Springer, London, 2013. 73-123.
10. CNIL's PIA Tool (https://oecd-opsi.org/innovations/pia-tool/#:~:text=The%20PIA%20tool%20is%20a,demonstrate%20compliance%20with%20the%20GDPR.),25 June, 2019
11. FUJITSU (2010) Personal data in the cloud: a global survey of consumer attitudes. [WWW document] http:/ /www.fujitsu.com/global/news/ publications/pataprivacy.html (accessed 20 March 2012).
12. Troncoso, Carmela. "PRIVACY & ONLINE RIGHTS KNOWLEDGE AREA." (2019).