Electronics and Computer Science
Faculty of Physical Sciences and Engineering
University of Southampton

Mohammad Ali Khan

29th April 2014

PrivacyMatters – resourceful privacy policy visualisations of UK/EU companies

Project supervisor: Dr. David Millard
Second examiner: Dr. Markus Brede

A project report submitted for the award of
MEng Computer Science

Abstract

Companies provide users with privacy policies that explain how their information is stored, yet these policies are filled with legal detail that renders them nigh incomprehensible. In this paper we propose an alternate solution by utilising the government-provided public data controller registry; this enables us to extract details about all the data controllers and their practices regarding data collection, which we can then display to users in an understandable and visually appealing form. We give some background information about these registries, discussing the merits of using a modern better-performing storage solution such as NoSQL and the appeal of having our solution displaying open Linked Data properties. We follow the process of designing a system for such a solution and walk through the steps of its implementation. We finally summarise the steps taken to evaluate the success of this project, draw conclusions from our evaluation and comment on future work.

Contents

1 Introduction

This section introduces the problem at hand and discusses the different resources which help form a direction for our project.

1.1 Company Privacy Policies

Data is important. People who interact with many companies for a variety of utilities have to give out their information to such establishments. However, there is always a risk of having such information being misused or mishandled, such as online businesses selling user data to third parties or spamming users with unwanted emails (Miyazaki & Fernandez, 2000). To tackle this problem, companies have privacy policies available for users to go through and understand their data handling practices. Unfortunately, these policies are either filled with difficult legal terms (Milne & Culnan, 2004) or contain complex vocabulary and sentence structure requiring a college education to be read properly (Anton, et al., 2004); therefore, users tend to not read them. A detailed analysis was carried on a number of retail, news site, internet service provider and travel agent companies with questions asked regarding data collection, data storage, data sharing and third-party data collection (Pollach, 2007). This study showed that the company policies used clever mitigation, obfuscation, enhancement and omission to make their policies sound as ambiguous a possible. According to the author, this implied that such companies are more interested in being covered legally than to genuinely provide user with proof of fair data trading. She calls for presentation of different data types and their handling methods with a simpler explanation so that users are fully aware of how their data is being used.

1.2 Data Controller Registers

We can define the aforementioned companies as data controllers which decide what, why and how customer information is processed (Act, 1998). The European Data Protection Directive (Commission, 1995) aims to regulate processing of personal data within the European Union by requiring such data controllers to provide their national authority with the following information (Article 19):

- name and address of the controller and of their representative, if applicable;
- purpose(s) of processing;
- description of category/categories of data subject and related data;
- category/categories of recipients to whom the data might be disclosed;
- proposed transfers of data to third countries;
- description to allow assessment of the measures taken to ensure security of processing

This, coupled with Article 21.2 which requires the member state to provide a process operations register containing at least the information mentioned above and make it accessible to anyone, provides a legal obligation for companies to provide their privacy policy information which can then be used to inform the public better about how their data is used. Such information about the data controllers is stored in a public register which can be sifted through to obtain the required information.

For United Kingdom, the register is available online and allows the users to search for data controllers and view them. However, the website is not user-friendly. It has a static basic structure with no indication or explanation about the different terms used. There is also a lack of analysis of the data controllers and interlinking between them, thereby not allowing people to understand more about their information and the relations between different controllers. Fortunately, the Information Commissions Office provides the contents of the register upon request.

1.3 Project Direction/Objectives

As company privacy policies are obviously ambiguous and rarely read, there is room to create a simplified version of these policies, with better information grouping. Public data controller registers are able to provide the information contained in company privacy policies but their representation is insufficient. Instead, the registry they provide on request can be used to create a visually appealing interface to an advanced register and help users understand their handled data better. The given information can also be used to carry out further analyses and allow for statistical grouping of companies according to various criteria. The register can combine all of this with interlinking between different companies, and be displayed on an easily accessible web interface, becoming a valuable resource for the public. This website can also act as an open data source, paving the road for potential future development.

2 Literature Review

This section delves into the different areas of computer science covered by our proposed solution. It aims to provide background information to give us a better perspective of the problems at hand.

2.1 Privacy Policies and Data Control Registers

There has been work done regarding presentation of company privacy policies in a machine readable format. One of the earliest forays into this field was by Platform for Privacy Preferences Project (P3P) (Cranor, et al., 2002) .Development for this was started in 1997, by W3C and the intention was to provide privacy policies in a format which was readable by web browsers. The major components of the P3P specification were (Cranor, 2003):

- Entity – contact information for the business
- Access – can user know the information kept about them
- Disputes – how to resolve privacy-related issues with the site
- Data – information collected
- Purpose – how information is utilised and if opt in/out option exists
- Recipient – when can data be shared and if opt in/out option exists
- Retention – when is data removed
- Consequence – human-readable version of data policy

Using this would mean that a user could set privacy preferences for themselves which can be compared with a website that they visit. The privacy policy of that company could be in an XML format and easily retrievable by a web browser or other user agents. Such agents could then display information to the user or compare it with the user-set preferences and take action accordingly, such as notifying the user if conditions had not been met etc.

P3P was considered in the context of EU law but after careful consideration and evaluation, the language was concluded to be not rich enough to describe EU Data Protection Law (Fischer-Hübner, 2001). Moreover, P3P was not fully implemented or adopted by websites in general (Beatty, et al., 2007). Many reasons were cited for its inevitable failure, key ones being the complexity of the language and the fact that its policy files required companies to be more transparent than they wanted (Schwartz, 2009). Instead, public data controller registers were used as a source of data on policies as they already existed and had solid legal backing.

2.2 Storage and NoSQL

The public registry contains more than 370,000 data controllers. To make the information for each data controller available on the web, we would need to store such details in a structured manner. This could done using SQL by defining different tables and filling them with data controller details, combining all the information using queries to display on the screen. However, after having looked at the register files, this would be the wrong path to take. We would first need to define strict tables to contain specific information about the controller and related it with its id. This means that we would have to perform complicated and computationally expensive queries to display all of the information. Relational databases such as SQL were developed for a time when little storage space was available (Couchbase, 2013), discouraging the duplication of data. Today, insufficient storage is not a concern so a solution of decreased complexity is preferred.

A good alternative are NoSQL databases. These are all the database management systems which do not follow the relational database pattern. Instead, they allow for a more lenient structure, have better performance and are very scalable. Some types of NoSQL databases are:

- Key-value based: data stored in a key-value format. Key also used as an index.

- Document-oriented: stores documents as are, indexing them and providing some querying mechanism.
- Extensible Record stores: allows partitioning of data vertically and horizontally across its nodes.

As there exist countless implementations of NoSQL databases, the user can choose the one which fits their requirements best (Cattel, 2011). Key-value stores are generally most suitable if we have just one type of object and need to search our data based on one trait. Document-oriented would work best when we have different kinds of objects whose structure may change over time and require searching with different fields. Extensible record store is valuable in applications demanding high concurrency.

NoSQL databases are not without flaws (Cattel, 2011). The reason they are faster and scalable is because they sacrifice the ACID properties associated with relational databases. While they are able to display an 'eventually consistent' property, it is not ideal in an application requiring high concurrency. SQL databases have been around for the past 30 years, forming a tight network: this is still the concept to learn in most universities and if someone is familiar with one Relational Database Management System, they find it easy to switch to another RDBMS. That said, NoSQL databases offer strong arguments and with extensive development being done on them in recent years, they will only get better (Cattel, 2011).

2.3 Linked Data

It would be ideal for our data to be available openly so that others can use it for their own purposes etc. Most companies or organisations provide their data to be utilised in many different non-propriety formats such as XML, JSON and CSV (Bizer, 2009). This requires specific APIs for each company as their information handles differently. Instead of having many different formats and APIs, it would be better to have the information available in the Web which others can point to, in the form of Linked Data.
Linked Data is basically linking data from various sources about a certain piece of data. The data is machine-readable and can be linked to and from external sources. Some of the rules for publishing, now known as 'Linked Data principles' have been defined as follows (Bizer, et al., 2009):

- URIs as to name things
- Use of HTTP URIs to allow people to look up names
- Looking up a URI gives relevant information using RDF, SPARQL standards
- Links to other URIs to allow for discovery of more things

Uniform Resource Identifiers (URIs) (Masinter, et al., 2005) provide generic means to identify objects. In case of these objects using *http://*, HTTP (Fielding, et al., 1999) is used to dereference the URI and look up the object, thereby providing neat, universal method for retrieving required information which is serialisable. Lastly, RDF provides a graph-based data model to structure and link the data of objects. It expresses the data as a subject, predicate and object triples. Out of these, the subject and object are both URIs or a URI and string whose relation is determined by a predicate, also in a URI form. Using these, Linked Data is able to add to general web architecture (Jacobs & Walsh, 2004), becoming another extra layer on top while maintaining a close connection. It is disjointed from the visual aspect of the web-pages, self-describing, open and has standardised access methods.

One of the desired end results of Linked Data is epitomised by the Linking Open Data Project (Bizer, et al., 2009). This was a community movement started by W3C Semantic Web Education and Outreach Group and the aim was to convert new and existing data available under open licences to an RDF format and publishing them online. Initially, this movement

was mainly headed by researchers and developers but with the passage of time, significant effort has been put in by some large organisations as BBC, Thomas Reuters and Library of Congress (Bizer, 2009). A visual representation of all the work done and Linked Data formed is given by the following figure.

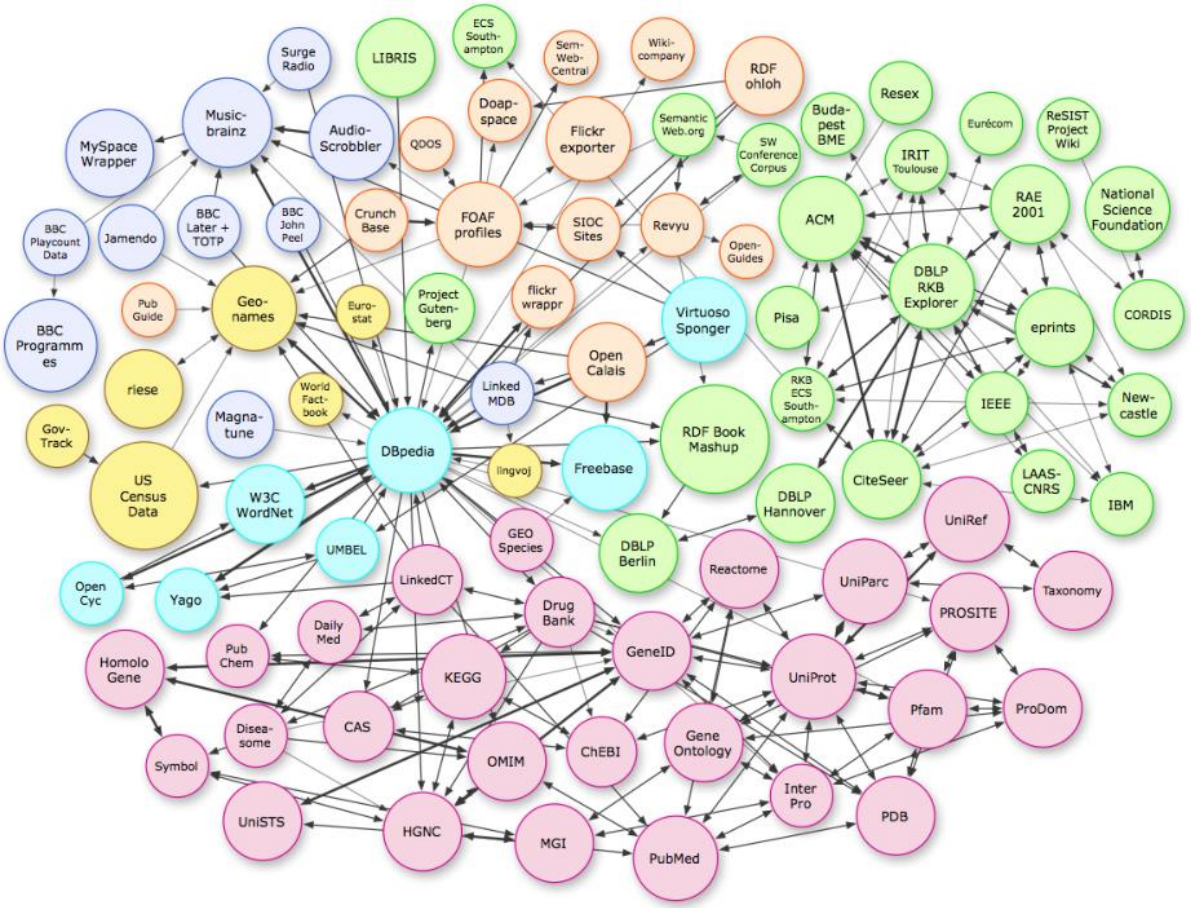

Figure 1. Visual representation of Linked Data (Bizer, et al., 2009)

For Open Data, a 5 star scheme was suggested by Tim-Berners Lee (Berners-Lee, 2006). The classification of the different stars is:

- 1 Star: Data available in any format on the web under an open license
- 2 Star: Data available in a structured, machine-readable format (e.g. as an Excel table)
- 3 Star: Data available in a non-propriety format (CSV, XML, JSON etc.)
- 4 Star: Standards such as RDF and SPARQL used to identify entities, allowing others to point to data
- 5 Star: Data also links to other data for more information

The stars can be displayed on each person's website to indicate their standard of open data to allow others work with it accordingly.

3 Specification

Before starting design or implementation of the system, we need to specify whom we are targeting with this project. We must also discuss the already existing technologies in this field and their shortcomings. Keeping those shortcomings and target audience in mind, we can draft a number of requirement necessary for the eventual evaluation before making technology and methodology decisions for our project.

3.1 Stakeholders

The basic stakeholders are the people who interact with any data controllers in our registry. These might be the data subjects whose information is collected by a data controller or a data disclose whom the data controller shares information with. It might be someone who wants to check a company they want to interact with in the future to see how their data would be handled. They might also just be comparing the difference in data processing details between different companies. As these people would range between different age groups, come from different back grounds and may or may not an understanding of privacy policies from before, we would need to use simple terminology provide good explanations where this is not possible. For those who want to analyse the data controller, we could provide additional information such as an indication of appropriateness regarding the amount of information collected. They might also want to see the popularity of a data information item e.g. data class, data subject etc. Lastly, they might want to see similar data controllers to different data information items. This information would also need to be provided in a format that does not clutter up the display. From a developer's point of view, there is room for this information to be available in the form of linked open data which can be referenced and used to perform further analyses, leaving room for improvements and creativity.

3.2 Existing Technology

There already exists an online version of the data controller register. It is hosted and maintained by the Information Commissioner's Office (ICO) and has a splash page with a text fields to allow the user to search for the data controller.

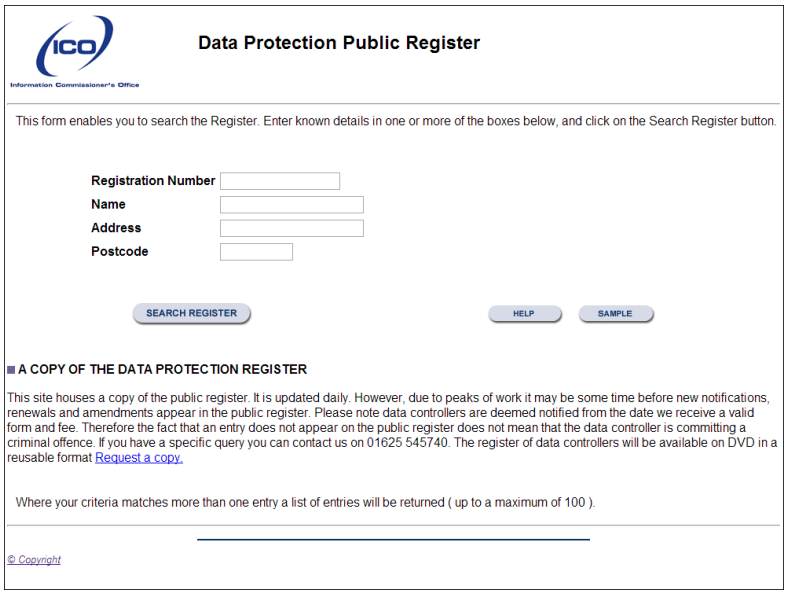

Figure 2. ICO home page

From this page, the user can search using the data controller name, registration number, address or postcode. If only one data controller is found pertaining to the search criteria, it is shown. Otherwise, a list of data controllers is provided but only if there are 100 or less data controller found for the query.

Once a data controller is selected, it is displayed for the user to sift through the different information.

Registration Number: Z6801020

Date Registered: 21 June 2002 **Registration Expires:** 20 June 2014

Data Controller: UNIVERSITY OF SOUTHAMPTON

Address:

HIGHFIELD
SOUTHAMPTON
HAMPSHIRE
SO17 1BJ

This data controller states that it is a public authority under the Freedom of Information Act 2000 or a Scottish public authority under the Freedom of Information (Scotland) Act 2002

This register entry describes, in very general terms, the personal data being processed by:

UNIVERSITY OF SOUTHAMPTON

Nature of work - University

Description of processing

The following is a broad description of the way this organisation/data controller processes personal information. To understand how your own personal information is processed you may need to refer to any personal communications you have received, check any privacy notices the organisation has provided or contact the organisation to ask about your personal circumstances.

Reasons/purposes for processing information

We process personal information to enable us to provide education and support services to our students and staff; advertising and promoting the university and the services we offer; publication of the university magazine and alumni relations, undertaking research and fundraising; managing our accounts and records and providing commercial activities to our clients. We also process personal information for the use of CCTV systems to monitor and collect visual images for the purposes of security and the prevention and detection of crime.

Type/classes of information processed

We process information relevant to the above reasons/purposes. This may include:

- personal details
- family details
- lifestyle and social circumstances
- education details and student records
- education and employment details
- financial details
- disciplinary and attendance records
- vetting checks;

Figure 3. Data controller page

As we can see from this page, this website is as basic as can be. There is little formatting such as the headings. The different lists of purposes, data classes etc. are fairly explained but that is all. There are no links to any other data controllers or any sort of analysis on the different data collected and processed by the data controller. Navigation-wise, there are no links allowing the user to move back to the search page or to search for another data controller from this page. Lastly, it is not possible for others to link to this data as there is no obligation for the register to work under an open license. These factors mean that the quality of the currently available technology is not good and there is room for a more useable product.

3.3 Requirements

Considering the short comings of the ICO website and the stakeholders, there is one major group of people to tend to while also providing basic support for the other group. The latter is not as necessary, only useful to have for future. There is a need of an interactive online platform which can provide data controller information in a structured format. There is also great need for simplicity and user-friendliness. We should aim to offer further analyses for data controllers and link to other data controllers to give a better representations for data controllers.

The requirements for this project are:

- Provide data controller details to user
- Provide information in a structured format
- Have a simple user interface
- Allow users to view data processing details
- Allow users to search for data controllers
- Link to external resources for increased richness of data
- Provide statistics for data processing details
- Provide useful visualisations for data controllers
- Provide all information on a single page
- Provide links to similar data controllers for data information items
- Provide a structured format for developers to point to
- Follow good user experience practices
- Have helpful explanations for many terms

3.4 Description of Proposed Solution

This project should result in an easily accessible web platform for users to search through the data controller register. Users should be able to search for a specific data controller easily and view its page. The page should show data controller details in a well-structured format, allowing users to click on different details and view statistics on the data and similar data controllers. There must be links to external pages for the data controllers and a structured format of this data controller obtainable and usable by other developers.

3.5 Decisions

3.5.1 Programming Languages

For parsing the XML data, Java will be used. This is the language we are most experienced in, having developed in it for three years. The backend of the website will also be implemented in Java, but there will be a need to learn about this as we have no prior experience in this.

For the web, HTML and CSS will be the technologies used. This is another field we have considerable experience in, having developed in them for the past three years. Once the layout and theme of the website has been finalised, implementing them will not be a problem.

JavaScript is a client-side scripting language run on the user's browser after loading of the page. This will be useful in loading different features in to allow for a smoother experience and reducing initial representation of data. We do not have much experience in this field, having worked with this language for half a year but there are countless resources available on the internet to make for a smooth sailing. JQuery, a library to use for JavaScript will also be

employed to make the code less verbose. Lastly, Moris.js, a library for neat charts, will be used to build statistics.

3.5.2 Storage

As discussed before, using a MySQL database for our project is not a good idea. Instead, the NoSQL database MongoDB will be used. This database is developer-friendly and allows us to be up and running quickly. Data is stored as documents in a JSON format, so there is no need for a fixed structure, a feature extremely useful to us. There is a Java Driver library which allows us to interact with the database from our Java program while parsing our data, allowing us to fill the data as we go.

3.5.3 Frameworks

The Twitter Bootstrap framework will be used for our website. This is a popular framework as it offers a standardised design structure, many easily usable features and great online support. This means that much load in terms of web designing is taken off us and handled by the framework.

For the backend of the website, the Play Framework will be used. This is a Java framework well-known for its highly dynamic structure. It has easy-to-use and powerful features, allowing us to save time in designing the back end by taking care of the whole matter for us. It utilises a templating engine which saves writing lots of static HTML code and allows for automatic code replication.

3.5.4 Tools

For version control, Git is used. This is a powerful version control tool which allows us to track our progress easily and also revert changes easily if things go wrong. Branching allows us to work on different aspects of the program simultaneously.

For programming, we make use of Eclipse, an IDE for our programming languages in use. For viewing of the XML files and occasional source files, Sublime Text 2 will also be used Google Chrome will be used for viewing and debugging our website.

3.5.5 Software Methodology

We aim to work iteratively on this project. Initially, we aim to divide this project into three main iterations. The first iteration will require a basic form of the system to be up and running. This will heavily depend on getting the parsing to work properly and building the database. The second iteration will deal with improving on the currently available display and adding analyses to it, beautifying the website etc. The last iteration will involve further complicated analyses, such as comparing two data controllers or selecting and comparing a bunch of data controllers. After each iteration, requirements and design may be reviewed.

4 Design

We must now work on the structure of our system. We need to determine all the possible uses for our system, look at the data we will be dealing with and finalise a plan to work with it. Lastly, we must come up with some concepts of how we want our website to look.

4.1 Use Cases

We can come up with numerous use cases for our system. Generally, the system has to allow the user to search for and view the data controller while providing a structural format and some useful statistics. There is a need to link to other data controllers in terms of relevance functionality to access the semi-structured, machine-readable form of the data for the developer. At the end of these requests, the system will retrieve the details from the database for the browser to display.

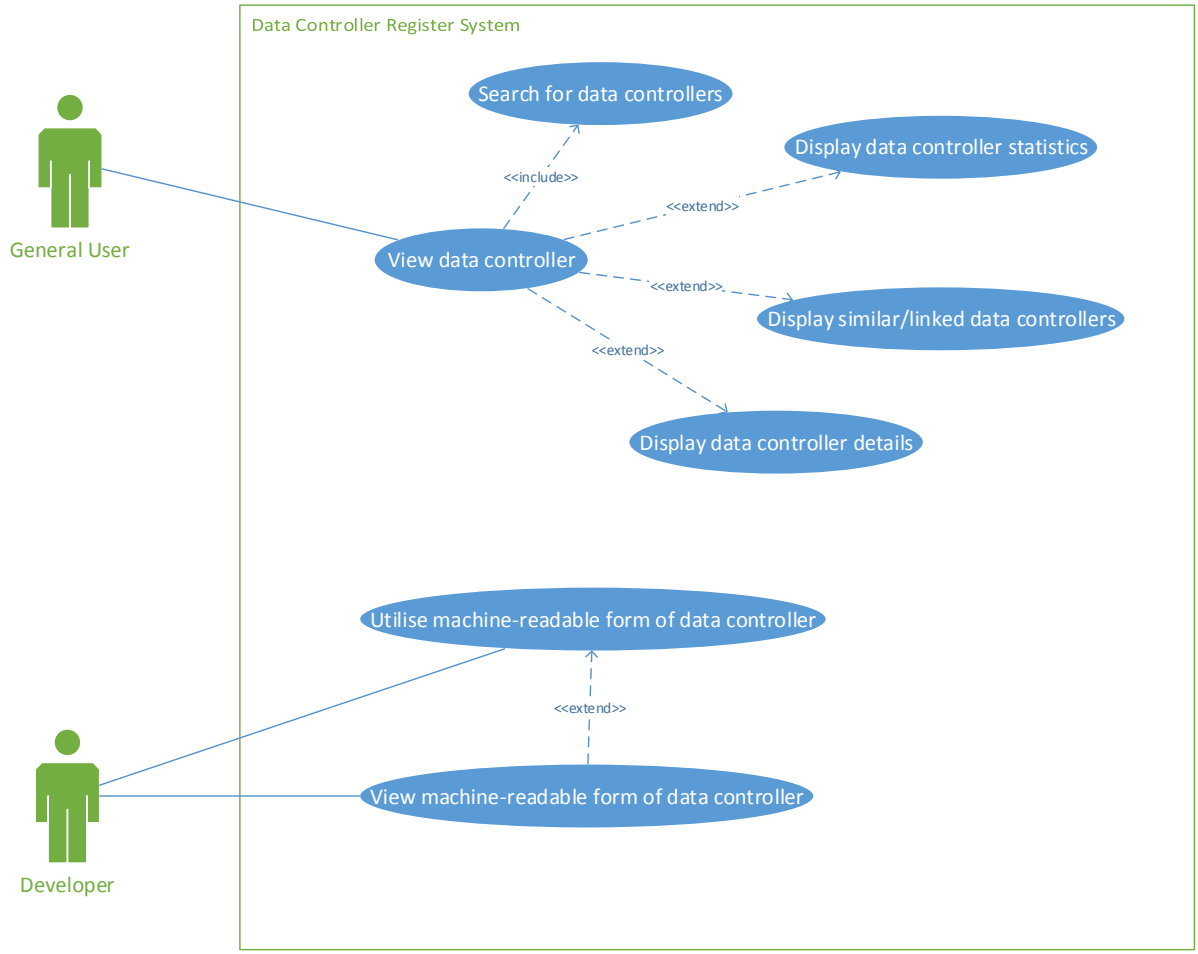

Figure 4. Use Case Diagram

4.2 Model Design

When requested, the ICO office provides a copy of the registry in a DVD. This DVD consists of an XML file containing the registry. In case of the registers of 2011 and before, the data was available in a very well-formed XML format. A record of a data controller would be defined as follows:

```
1.  <DATA_CTLR_DETAILS>
2.    <DATA_CTLR_NAME>INTEGRITY WILLS (UK)</DATA_CTLR_NAME>
3.    <ADD_1>STERLING COURT</ADD_1>
```

```
    <ADD_2>4 GRESHAM ROAD</ADD_2>
    <ADD_3>BRENTWOOD</ADD_3>
    <ADD_4>ESSEX</ADD_4>
    <PCODE>CM14 4HN</PCODE>
    <REG_NO>Z9146121</REG_NO>
    <DATE_REGISTERED>2005-08-01</DATE_REGISTERED>
    <DATE_EXPIRES>2011-07-31</DATE_EXPIRES>
    <FOI_MARKER>N</FOI_MARKER>
    <VOLUNTARY>N</VOLUNTARY>
    <EXEMPT_PROCESSING>N</EXEMPT_PROCESSING>
    <OTHER_NAMES></OTHER_NAMES>
    <PURPOSES>
      <PURPOSE>
        <PURPOSE>Staff Administration</PURPOSE>
        <OTHER_PURPOSE></OTHER_PURPOSE>
        <SUBJECTS>
          <SUBJECT>Staff including temporary and casual workers</SUBJECT>
          <SUBJECT>Guardians and associates of the data subject</SUBJECT>
        </SUBJECTS>
        <CLASSES>
          <CLASS>Personal Details</CLASS>
          <CLASS>Family, Lifestyle and Social Circumstances</CLASS>
        </CLASSES>
        <RECIPIENTS>
          <RECIPIENT>Data subjects themselves</RECIPIENT>
          <RECIPIENT>Current, past or prospective employers of data subject</RECIPIENT>
        </RECIPIENTS>
        <TRANSFERS>
          <TRANSFER>None outside the European Economic Area</TRANSFER>
        </TRANSFERS>
      </PURPOSE>
    </PURPOSES>
</DATA_CTLR_DETAILS>
```

Listing 1. Pre-2011 data controller XML record

This is 3-star data as it is well-formed and available in a machine-readable format. Information is given in a descriptive format, making it easy for a programmer to sift through.

However, the latest registry copies have changed their data format. The nice list of purpose tags has been replace by one tag: `<Nature_of_Work_description>`. This tag contains all the data processing details.

```
  <Registration>
    <Record>
      <Registration_number>ZA013235</Registration_number>
      <Organisation_name>ERESBY SPECIAL SCHOOL</Organisation_name>
      <Organisation_address_line_1>ERESBY AVENUE</Organisation_address_line_1>
      <Organisation_address_line_4>SPILSBY</Organisation_address_line_4>
      <Organisation_address_line_5>LINCOLNSHIRE</Organisation_address_line_5>
      <Organisation_postcode>PE23 5HU</Organisation_postcode>
      <Organisation_country>UNITED KINGDOM</Organisation_country>
      <Freedom_of_Information_flag>Y</Freedom_of_Information_flag>
      <Start_date_of_registration>2013-06-14</Start_date_of_registration>
      <End_date_of_registration>2014-06-13</End_date_of_registration>
      <Exempt_processing_flag>N</Exempt_processing_flag>
      <Contact_in_UK_C1>None</Contact_in_UK_C1>
      <Subject_access_Contact_C2>None</Subject_access_Contact_C2>
      <Nature_of_Work_description></Nature_of_Work_description>
    </Record>
  </Registration>
```

Listing 2. Post 2011 data controller XML record

There has also been an emergence of a new format for representing information. As seen before, we had separate details for each purpose in separate data classes, subjects etc. This has been changed to have a singular set of data purposes and a set of data classes, subjects and discloses each. It is not known anymore which data class or subject belongs to which purpose. This change has decreased the richness of our data but two new features in Nature of Work of the data controller and sensitive data classes has allowed us to categorise data controllers better and add a further dimension to our information.

In the light of this, the data controller details are divided into two formats: the new format and the old one. The old format is a list of purposes, each containing data classes, data subjects etc. while the new format has one list of purposes and a generic list of all the other details. With this in mind, we came up with a DataController class, which consisits of the common information contained in the two formats. This class may have a NewFormat object, or a List of Purposes. We must also remember that we are using a document-oriented NoSQL database and hence our classes can have any sort of structure.

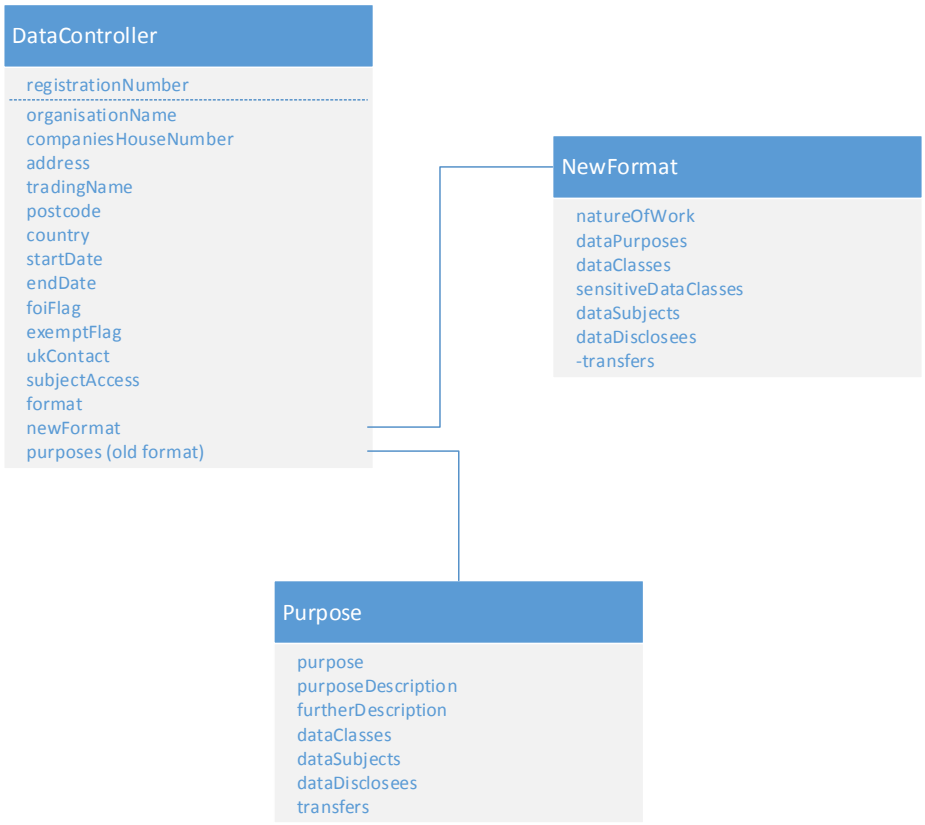

Figure 5. Class diagrams for our data controller models

Field Name	Type	Description
registrationNumber	Eight character String	Identification number for each data controller
organisationName	String	Name of the data controller
companiseHouseNumber	String	Companies House number , if exists
tradingName	String	Trading name of data controller, if exists
address	Array of Strings	Array containing lines of data controller address
postcode	String	Postcode for data controller
country	String	Data controller country
startDate	Date	Start date of registration with data controller register
endDate	Date	End date of registration with data controller register
foiFlag	String	Flag to determine if data controller is a public authority or not
exemptFlag	String	Flag to determine if data controller is exempt from informing register of some of the data it processes
format	String	Format of the data processing details
newFormat	NewFormat class	Class for new format of data processing details
purposes	Array of Purpose objects	List of purpose pertaining to the old format of data processing details
natureOfWork	String	Determines the type of data controller
dataPurposes	Array of Strings	List of purposes for collecting data
dataClasses	Array of Strings	List of information collected
sensitiveDataClasses	Array of Strings	List of sensitive information collected
dataSubjects	Array of Strings	List of people information is collected from
dataDisclosees	Array of Strings	List of people information may be to disclosed to
transfers	String	Statement informing about the transfer policy of the data collected
purpose	String	Name of purpose for collecting data
purposeDescription	String	Description of the purpose
furtherDescription	String	Further description of the purpose, if added by the data controller

Table 1. Data dictionary for our data controller models

We also want to keep links between different data controllers and provide useful statistics and visualisations. There is no need to build all of this dynamically with different queries for querying our huge database in real-time will result in a slow performance. Moreover, our database will always be static, unless we are rebuilding it with a new register file. Therefore, it makes sense to pre-process our data and build up all our tables so that in real-time, we just fetch the different values. This means we run our first program to build our database. We can then run another program, which sifts through the data controller register, building statistics from it. We can have a class which stores the type of information and all the data controllers related to it for linking. This will have a record for each data controller details such as purposes,

nature of work, data classes, data subjects etc. For example, for a data class ‘personal details’, we will this as the type of record and all the controllers related to it. For purposes and nature of work, we must take more information such as medians for the number of data classes, subjects and disclosees listed. These result in three classes, which make efficient use of inheritance. We also use another class, RegistryListItem, which is a small class only to hold the registrationNumber and controller name for identification in the database. We have one other class to collect general statistics and information on the data controller register.

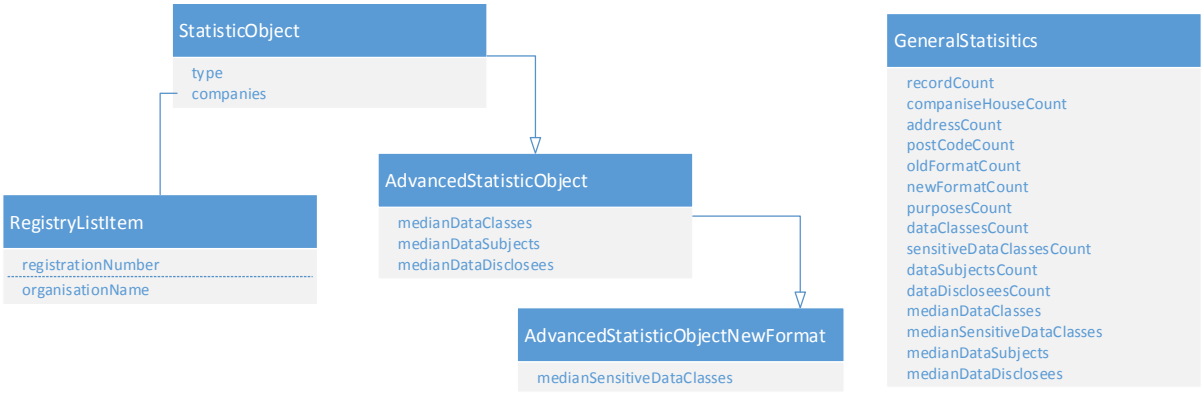

Figure 6. Models for our statistics

Field	Type	Description
type	String	Identifier for data record belongs to. Will be a member of one of data processing detail lists
companies	Array of RegistryListItem objects	List of companies sharing that item
registrationNumber	Eight character String	Identification number for each data controller
organisationName	String	Name of the data controller
medianDataClasses	Integer	Median amount of information taken
medianDataSubjects	Integer	Median number of data subjects information is taken from
medianDataDisclosees	Integer	Median number of people information is disclosed to
medianSensitiveDataClasses	Integer	Median number of sensitive information taken
recordCount	Integer	Number of records in register
companiesHouseCount	Integer	Number of data controllers with a Companies House Number
addressCount	Integer	Number of data controllers with address given
postCodeCount	Integer	Number of data controllers with postcode given
newFormatCount	Integer	Number of data controllers with new format of data processing details

oldFormatCount	Integer	Number of data controllers with old format of data processing details
purposeCount	Integer	Total number of different purposes cited
dataClassesCount	Integer	Total number of data classes collected
sensitiveDataClassesCount	Integer	Total number of sensitive data classes collected
dataSubjectsCount	Integer	Total number of data subjects collected from
dataDiscloseesCount	Integer	Total number of data disclosees disclosed to

Table 2. Data dictionary for statistics models

4.3 Architecture

The system design is simple. We take in data from the register XML file and parse through it. We build an entry for each data controller and add it to our database. The data will remain static so this is a one-off process, happening only to load the new registry file. The database will then interact with the server, with the server retrieving lists of data controllers and other information as well as querying for specific information. This will then be sent to the client-side in a JSON format and manipulated accordingly by it to display in the required format and layout with the help of JavaScript, HTML and CSS.

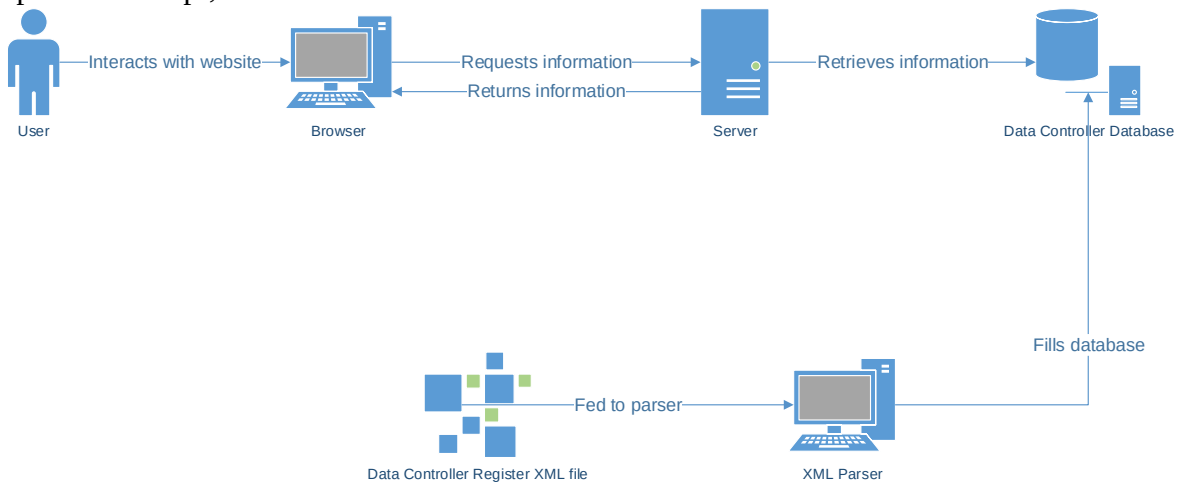

Figure 7. Conceptual diagram showing an overview of the system

The parsing and building of the database is disjoint from the website. The parsing program cleans the database and builds it up every time a new file added. It also iterates through the database, pre-processing the data by building statistics of and links between data controllers.

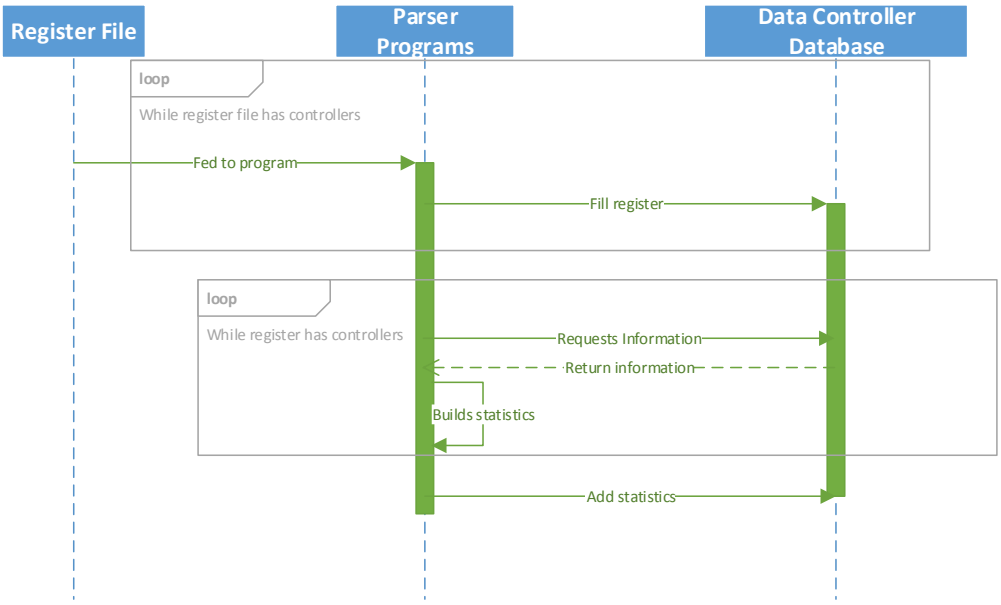

Figure 8. Sequence diagram for parser

On the web platform, all of the data is sent to the client browser at once and it can then build the different visualisations at request of the user. This means no real-time queries are made for data processing and requests are fulfilled quickly.

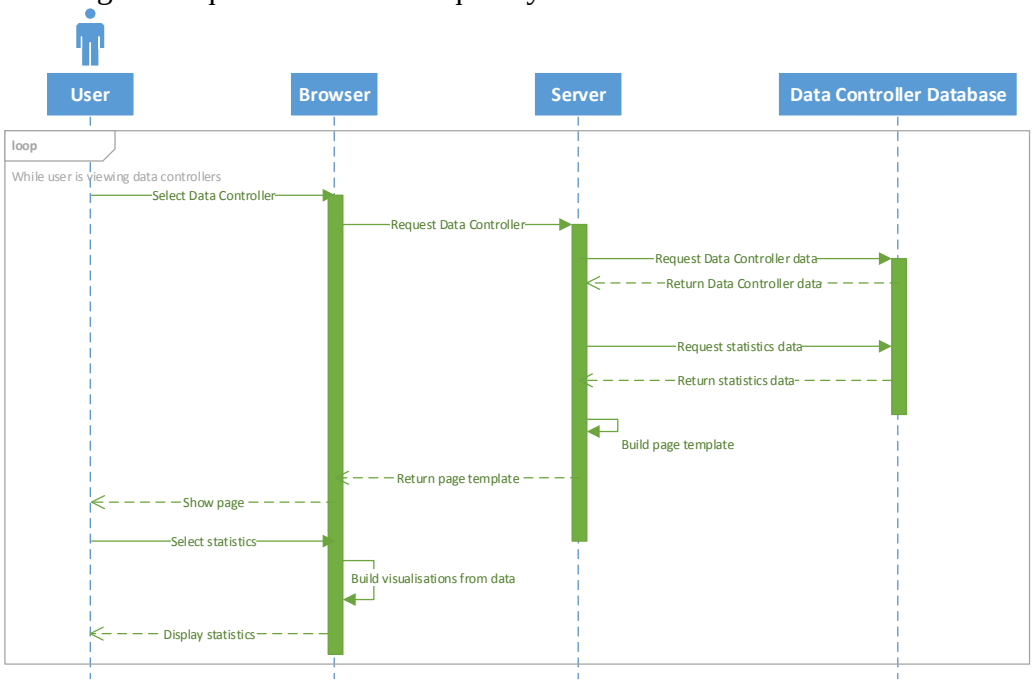

Figure 9. Sequence diagram for website

4.4 Wireframes

The biggest part of the project is related to visual representation of the data controller and the website in general. This is why we have aimed for utmost simplicity in our designs for our website. The home page is as simple as possible, providing no distractions to the user and allowing them to search for the required data controller.

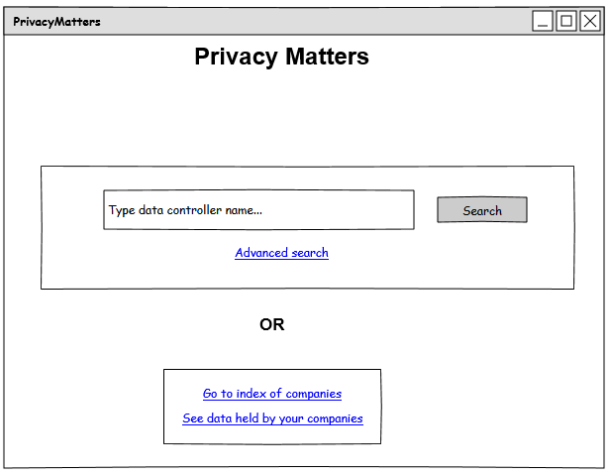

Figure 10. Wireframe for website home page

The data controller page must be divided depending on the information we have on it. We decided to have two modules of information at the top of the page, one to contain the general information on the data controller and the other to show contact information. The contact panel will also point the user to the location of the data controller on a Google Maps block. For data processing details, we provide a modular representation while saving as much space as possible. We have three boxes containing the data classes, data subjects and data disclosees respectively. This allows us to have a clean interface without having to scroll up and down a lot to view the information. We also make these list items clickable and allow statistics to pop up whenever we need them to.

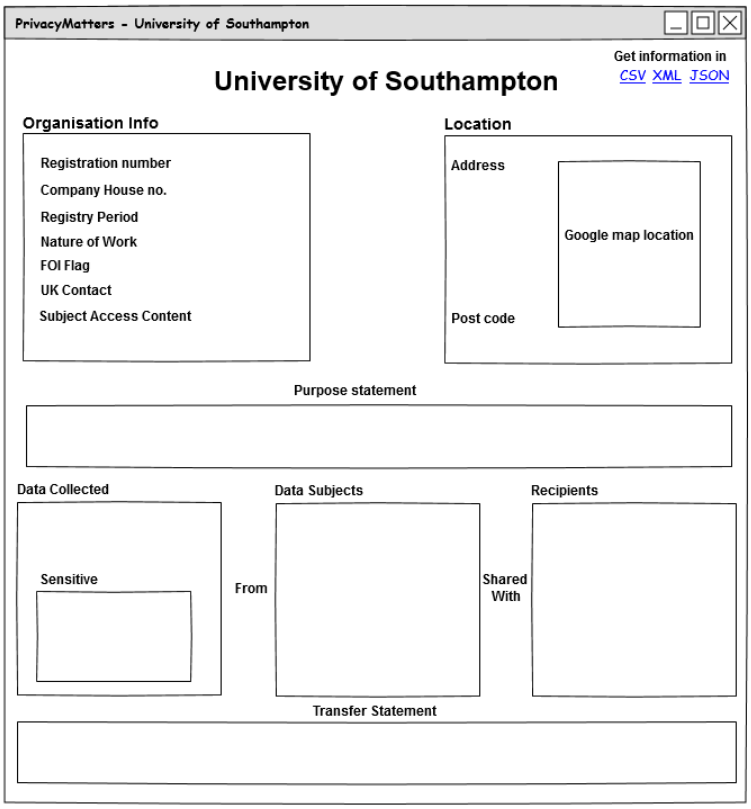

Figure 11. Data controller page wireframe

5 Implementation

This sections documents the path taken to build our system. It covers some of the underlying factors which influenced our decisions and shows the evolution of our ideas while making the best system possible.

5.1 Prototyping

5.1.1 Database

We wanted to prototype with different databases before deciding on a solution. We wanted a document-oriented NoSQL database because it would not reqiure a set structure, as desperately needed by our changeable format. After thorough research, we experimented with two databases: Couchbase Server and MongoDB. Both were document-oriented and each of them had their own advantages; MongoDB was more developer-friendly while Couchbase Server scaled better. Each database system was installed onto our machine and we attempted to implement a simple application. Couchbase Server was found troublesome to work with as there were problems with its installation and it was found difficult to understand and work with. In comparison, MongoDB installed with ease and worked perfectly in the experiments. It had driver libraries for Java and Python, which we downloaded and used. Using Python, we created a simple web form which allowed the user to create a new record for a guest, submitting a name and email. This was added to a MongoDB collection and displayed on the webpage simultaneously. With the Java Driver, we implemented a class to act as a handler for MongoDB, containing methods to create a database or collection and work with records. We also ran small demos with it, creating various databases, collections, and records. Satisfied, we decided to use MongoDB for our project.

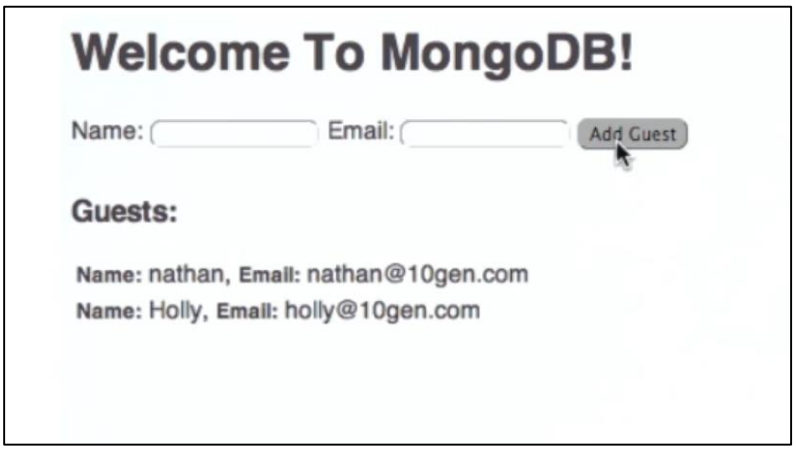

Figure 12. Guest list prototype with MongoDB

5.1.2 Play Framework

Before we started our implementation, we wanted to make a small-scale implementation of our project structure using our selected framework. This meant using our framework to work with a specific class having many attributes which we want to display on a separate page. This page may be reached from a list or directly by entering the unique id of the item on the address page. Consequently, we created a small Person class, containing name, age, date of birth and an id. We also made an array of Person objects and were successfully able to display them in a list. Our controller class would handle all the requests made for the different routes. Going to specific routes would trigger different methods which would be return different pages. When

the user went the home page, they automatically got redirected to the list of Person objects which would be available on localhost:9000/people.

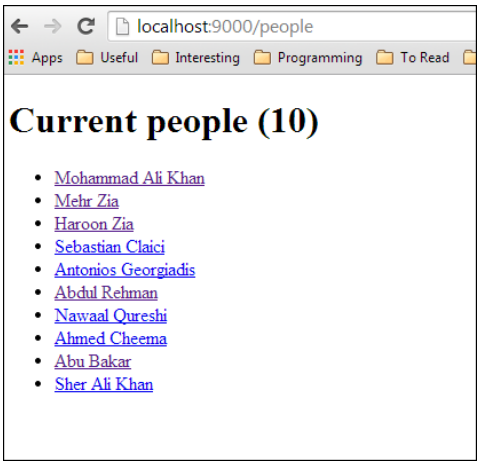

Figure 13. List of people

Once a person link was clicked, a request was made to the controller along with the id of the person clicked on. This person was then retrieved from the array and returned to the page, where the templating engine was used to display the information in the desired manner. The user would be taken to localhost:9000/person/(id). The user can also just use the id of a person to reach the page quickly or link it to someone easily. Once this page was reached person details were displayed.

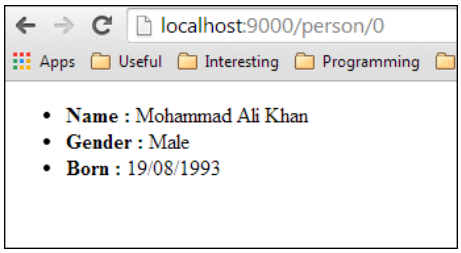

Figure 14. Individual person page

5.1.3 Charts

We attempted to implement different charts for our statistics. The purpose of this was understanding how to make it work for our project and experimenting with different JavaScript chart libraries. Ee made a small webpage, in which we add different hidden figures. This was done to mirror the future functioning of our system, which would return hidden values to be used to make charts at the user's request. We experimented with three different JavaScript libraries: d3.js, charts.js and Moris.js. With each, we created a small chart with the help of hidden values. We also tried to make them appear when a button or link was clicked. Out the three, d3.js was found to be the most complicated and overpowered library; it offered a vast number of features but we required something simpler and easily implementable. Charts.js and Moris.js fell into this category but charts.js did not scale well dynamically; they required decent amounts of space to be displayed properly while Moris.js would rescale extremely. Therefore, we decided to use Moris.js for our charts in our project.

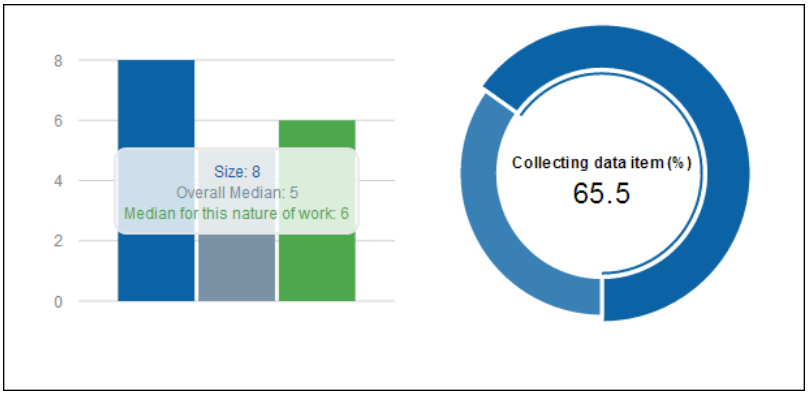

Figure 15. Sample Moris.js charts

5.2 First Iteration

We started the first iteration with the aim of having a basic website up. This meant developing the parser, adding records to our database and making sure each data controller was viewable on the website in the expected way.

5.2.1 Parsing

We started work on our parser in hope of finishing the building of our database quickly. However, this was not possible and we lost some time due to some problems with our data file.

As mentioned before, both data formats were present in the `<Nature_of_Work_description>` tag. A sample of each data type is given below.

```
<P>
    <FONT size=2 face=verdana><STRONG>Purpose 1</STRONG></FONT>
</P>
<P>
    <FONT size=2 face=verdana>Education</FONT>
</P>
<P>
    <FONT size=2 face=verdana><STRONG>Purpose Description:</STRONG></FONT>
</P>
<P>
    <FONT size=2 face=verdana>The provision of education or training
        as a primary function or as a business activity.</FONT>
</P>
<P>
    <FONT size=2 face=verdana><STRONG>Data Subjects are:</STRONG></FONT>
</P>
<P>
    <FONT size=2 face=verdana>Suppliers<br>Complainants,
        correspondents and enquirers</FONT>
</P>
<P>
    <FONT size=2 face=verdana><STRONG>Data Classes are:</STRONG></FONT>
</P>
<P>
    <FONT size=2 face=verdana>Personal Details<br>Family,
        Lifestyle and Social Circumstances<br></FONT>
</P>
<P>
    <FONT size=2 face=verdana><STRONG>Sources (S) and
            Disclosures (D)(1984 Act). Recipients(1998 Act):</STRONG></FONT>
</P>
<P>
    <FONT size=2 face=verdana><br>Data subjects themselves<br>Employees and
```

```
34.         agents of the data controller</FONT>
35. </P>
36. <P>
37.     <FONT size=2 face=verdana><STRONG>Transfers:</STRONG></FONT>
38. </P>
39. <P>
40.     <FONT size=2 face=verdana><br>None outside the European
41.         Economic Area</FONT>
42. </P>
```

Listing 3. Sample of old information format

The new format is also displayed similarly.

```
1.  <B><FONT size=2 face=verdana>
2.          <P>Nature of work - Academy</P>
3.          <P></P></B>
4.  <P>
5.      <B>Description of processing<BR></B>The following is a broad
6.      description of the way this organisation/data controller processes
7.      personal information. To understand how your own personal information
8.      is processed you may need to refer to any personal communications you
9.      have received, check any privacy notices the organisation has provided
10.     or contact the organisation to ask about your personal circumstances.
11. </P>
12. <P></P>
13. <P>
14.     <B>Reasons/purposes for processing information<BR></B>We process
15.     personal information to enable us to provide education, training,
16.     welfare and educational support services, to administer school
17.     property; maintaining our own accounts and records, undertake
18.     fundraising; support and manage our employees.
19. </P>
20. <P></P>
21. <P>
22.     <B>Type/classes of information processed</B><B><BR></B>We process
23.     information relevant to the above reasons/purposes. This may include:
24. </P>
25. <UL>
26.     <LI>personal details</LI>
27.     <LI>family details
28. </UL>
29. We also process sensitive classes of information that may include:
30. <UL>
31.     <LI>physical or mental health details
32.     <LI>racial or ethnic origin
33. </UL>
34. <P>
35.     <B>Who the information is processed about<BR></B>We process
36.     personal information about:
37. <UL>
38.     <LI>employees
39.     <LI>students and pupils
40. </UL>
41. <P>
42.     <B>Who the information may be shared with<BR></B> Where necessary
43.     or required we share information with:
44. <UL>
45.     <LI>financial organisations
46.     <LI>press and the media</LI>
47. </UL>
48. </FONT>
49. <B><FONT size=2 face=verdana>
50.         <P>
```

```
51.             <BR>
52.         </P>
53.         <P>Transfers</B>
54. </P>
55. <P>It may sometimes be necessary to transfer personal information
56.     overseas. When this is needed information is only shared within the
57.     European Economic Area (EEA).</P>
58. </FONT>
```

Listing 4. Sample of new data information format

In this data sample, the data has a certain pattern. It has headings in `<B>` or `<STRONG>` tags that we can expect. With this in mind, we tried to make use of a HTML parsing library to retrieve information from the data.

The use of this library was not helpful. While there was a pattern to the previous data format the new format was found to be continuously inconsistent and badly formed. The html parser library allowed us to categorise the different pieces of text according to different tags. Using the `<B>` tag to categorise the headings seemed a good idea but soon we discovered that all the headings were not encompassed within a `<B>` tag. It may sometimes be a `<STRONG>` tag or may be neither. We needed to check for all cases which made automation tedious. Moreover, all the headings were not present at which meant that we could make no assumptions about the data.

There was also inconsistency with respect to the data classes, subjects etc. Generally, the list of data classes, data subjects and discloses was given in an unordered list but on many occasions, data was given in a block of text. This would require us extract the terms from the prose was not guaranteed to be correct, meaning we would lose the richness of our data.

In the end, we decided another approach. Instead of using the HTML parser and treating the text as HTML, we just stripped out all the tags to give ourselves a list of strings. From this list, we found different headings, handling them accordingly. This was successful but there were many assumptions made on the data which, if found false, would break down the program. In the end, a best fit solution was found, which compensated for the different ordering of headings. When tried on 30 formats, this was successful.

There were bound to be cases where the format will not be consistent and the information for some data controllers may not be represented properly. However, the number of such mistakes would only be a small percentage of the overall data controllers. It must be taken into perspective that there are roughly 375,000 data controllers present in our register file and even achieving an 80% accuracy would be great, although the accuracy was expected be more than that. Therefore, we continued with this solution and finished the parsing aspect of our project, albeit not as quickly as we hoped.

5.2.2 Initial Deployment

Once we sorted out the parsing, we carried on with deploying our website. As we were using the Play framework for our website, we needed to find a platform which would be able to run our framework. We considered different platforms on which to host our application: Google AppEngine, CloudBees, Amazon Web Services and Heroku. Out of these, Google AppEngine was not compatible with the latest version of the Play framework, and Amazon Web Services required a WAR file .The best two options were found to be Heroku and Cloudbees but the Cloudbees interface for hosting and managing applications was found extremely confusing to

work with. In contrast, Heroku had a simple process for deploying and setting up a Play application and was therefore chosen to host our application.

We also needed to find an online storage for our MongoDB database. Our university did not have a Mongo database on their server so we explored other options. Finally, we registered with MongoLabs, which allowed us to have a database of maximum size 512MB for free. This service gave us a URI to connect to the database and interact with it, storing our data controllers in a JSON format. This suited us as we could work with a small prototype of our database quickly and without cost during development. We decided to work with a smaller number of data controller during the development of our project and if it was successful, we would explore options to store the whole register. In any case, we could always run the complete database locally if we wanted to.


```
{
    "_id": {
        "$oid": "535d2fb7d3c9f9d5aad51fce"
    },
    "registrationNumber": "ZA034589",
    "organisationName": "ANNEMARIE TIMONY",
    "companiesHouseNumber": "(none)",
    "postcode": "G11 6TG",
    "country": "United Kingdom",
    "foiFlag": "No",
    "exemptFlag": "No",
    "tradingName": "(none)",
    "format": "new",
    "startDate": {
        "year": 2014,
        "month": 0,
        "dayOfMonth": 3,
        "hourOfDay": 0,
        "minute": 0,
        "second": 0
    },
    "endDate": {
        "year": 2015,
        "month": 0,
        "dayOfMonth": 2,
        "hourOfDay": 0,
        "minute": 0,
        "second": 0
    },
    "address": [
        "Partick Dental Practice",
        "356 Dumbarton Road",
        "Glasgow"
    ],
    "newFormat": {
        "purposes": [
            "We process personal information to enable us to provide healthcare services to our patients and manage our employees."
        ],
        "dataSubjects": [
            "patients",
            "customers"
        ],
        "dataClasses": [
            "personal details",
            "family details"
        ],
        "sensitiveData": [
            "physical or mental health details",
            "racial or ethnic origin"
        ],
        "dataDisclosees": [
            "healthcare professionals",
            "social and welfare organisations"
        ],
        "natureOfWork": "Dentist",
        "transfers": "information is only shared within the european economic area (eea). any transfers made will be in full compliance with all aspects of the data protection act.",
        "otherPurposes": []
    }
}
```


Figure 16. JSON format of our data controller on MongoLabs

We then worked on our controllers, adding methods to redirect users to a registry page containing a number of data controllers. We retrieved our list of data controllers from the database and used a JSON library to retrieve different attributes from the JSON version of our controllers. We packed the controller name and registration number into a RegistryListItem object which could then be sent to the templating engine in an array. We could then show the controller names on the list and link to their address page with the registration number. Once we clicked on a data controller, we would be redirected to /datacontroller/(registration number). This was designed to allow for a unique address for each data controller and allowing the user

to reach the data controller page quickly if they already had the data controller registration number.

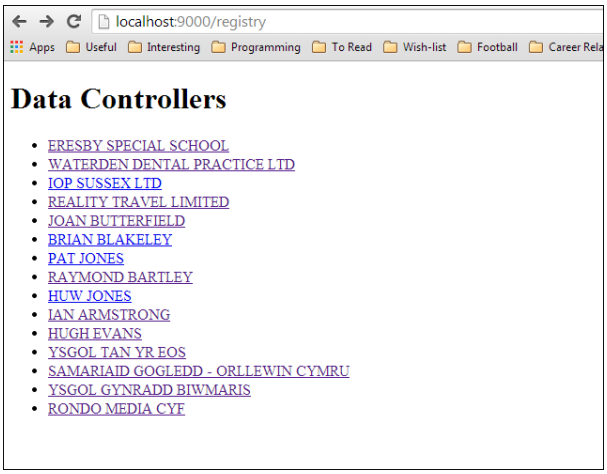

Figure 17. Initial representation of registry

When a request was made for a data controller, we searched for the data controller in our database, returning the JSON string. Once we received it, we used the gson library to unpack this string back to our DataController class and passed it on to our templating engine. Using the data controller object, we displayed the information from different attributes. For our initial representation, we had a basic version of data controller information on each page by grouping objects as per our design and presenting them in a <fieldset> tag.

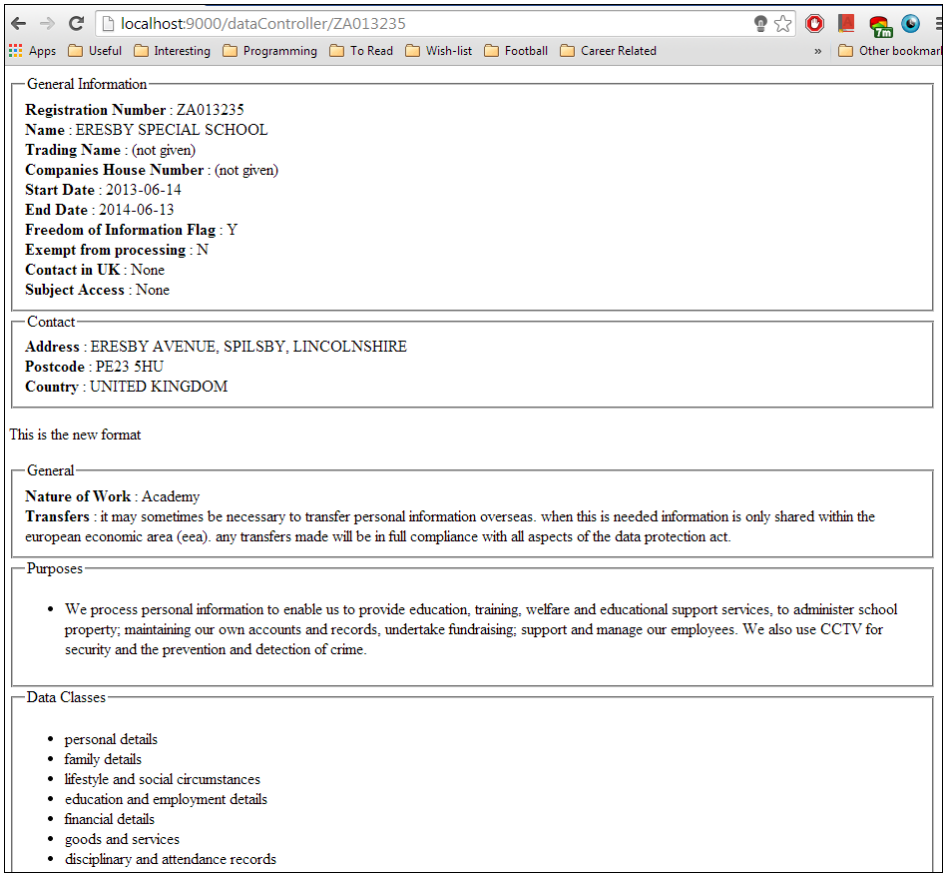

Figure 18. Initial data controller page

5.3 Second Iteration

5.3.1 Robust Parsing

After experiencing trouble with our parser at the beginning, we tried to test a greater number of data controllers to see how it fared. We ran into more trouble due to a few more assumptions on the data but these were easily cleared. Another thing to note about the newer format was that at times, they cited extra purposes for collecting data apart from the ones they had under "Reasons/purposes for processing data". After careful study of the available data and the new data controller registration form, we discovered that the register asked data controller to cite certain purposes separately. These were related to CCTV, consultation, trading and research aspects of the data controllers. These different purposes had their own headings in the new format and with our current parser, would filter through, being added to the previous information panel which had been detected. These purposes were written in prose, meaning there was not a way to filter our data classes, subjects and disclosees out of them. To compensate for this, we revised our model design, adding an extra class to add these purposes for the new formats. An updated design structure can be seen below.

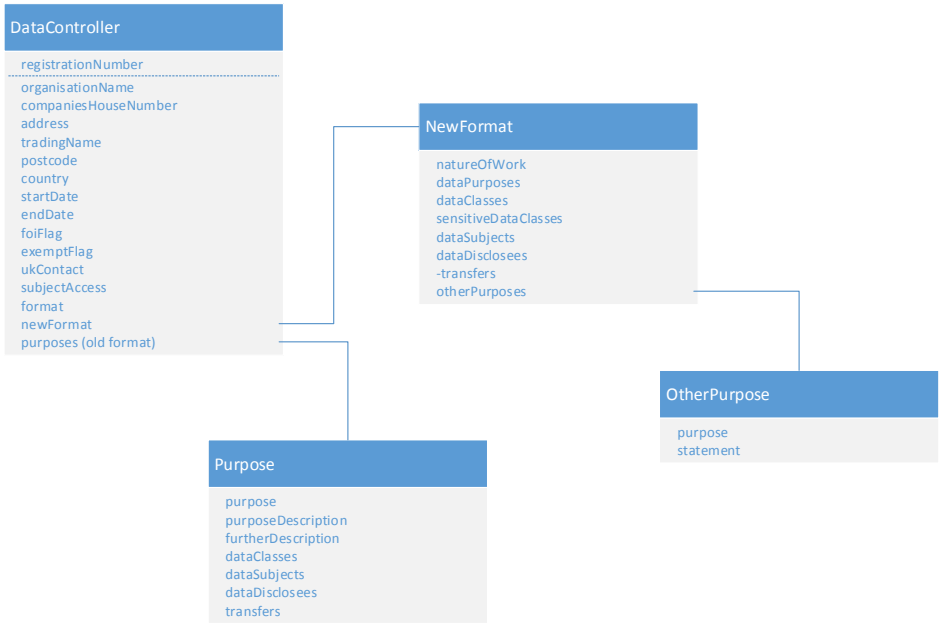

Figure 19. Revised data controller models

With our new headings added and lesser assumptions about the order of information appearing in our data, we made our parser more robust. This resulted in greater success in parsing our data and another aspect to our data.

5.3.2 User interface

Now that the foundation had been laid for our project, we started on improving the user interface. We made extensive use of the Twitter Bootstrap framework which did the heavy lifting for us, saving us immense amount of time.

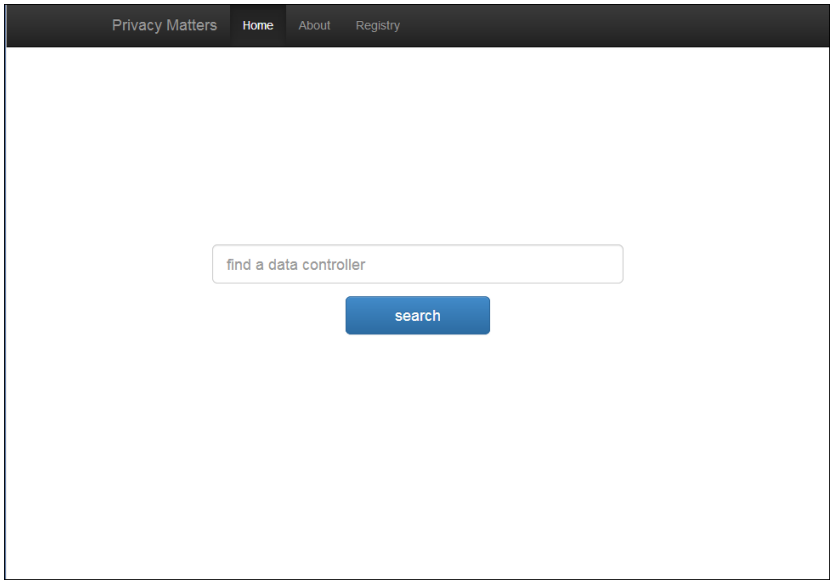

Figure 20. Home page

We started with our home page. We decided to make it even simpler by replacing all the other links from our page and having a search form in place. Search was not implemented currently so we had a link in the navigation bar to view the registry, a list of all the data controllers. Working with a small number of data controllers (100), this was possible. With the help of the grid system in bootstrap, we divided the top two boxes of information into General Information and Contact. In the contact panel, we had a canvas set up and JavaScript code to run. This code got the post code of the data controller from the page and centred their location on the map.

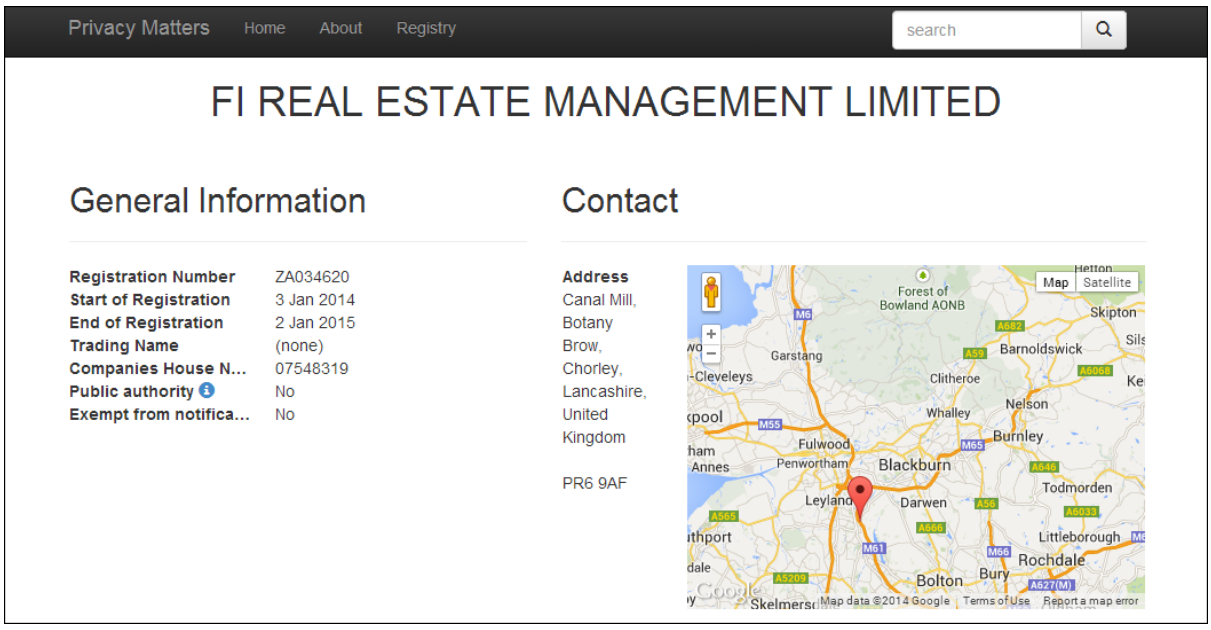

Figure 21. General information and Contact groups

Because of the changes realised in the data controller format, we also redid our design of the processing details. As both formats now had distinct purposes, we thought it would be better to have a similar pattern for each format. We also wanted an interface which was not as tedious to navigate through as the ICO pages, meaning not having to scroll up and down. Therefore,

we decided on clickable boxes of purposes. These would expand or close at the will of the user while taking away tedious scrolling.

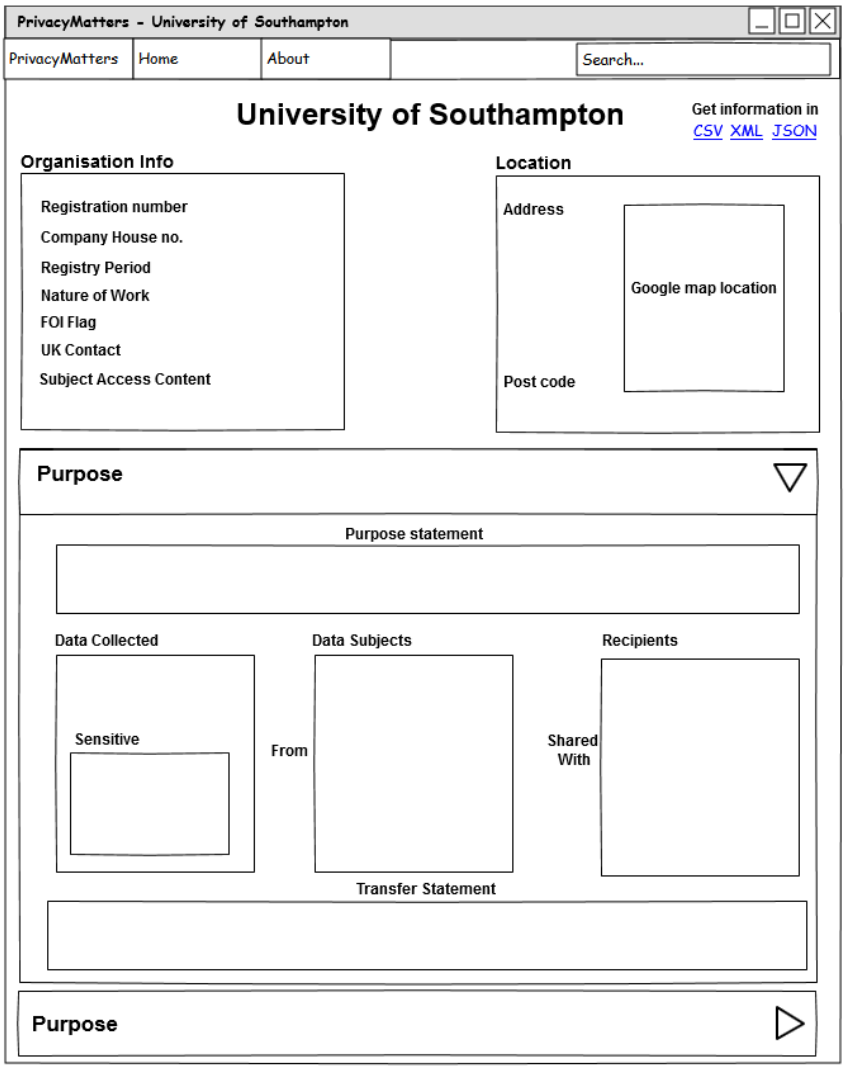

Figure 22. Revised data controller page design

We provided a page header titled “Data Processing Details” and had each purpose and their related information in a panel, with only the panel body visible. These boxes had icons indicating that they are expandable, displaying the other panels containing the list items of data classes, disclosees and subjects. For the newer format, the nature of work of the data controller was made visible at before the start of the data processing details. Finally, we were able to make our pages look cleaner and simpler.

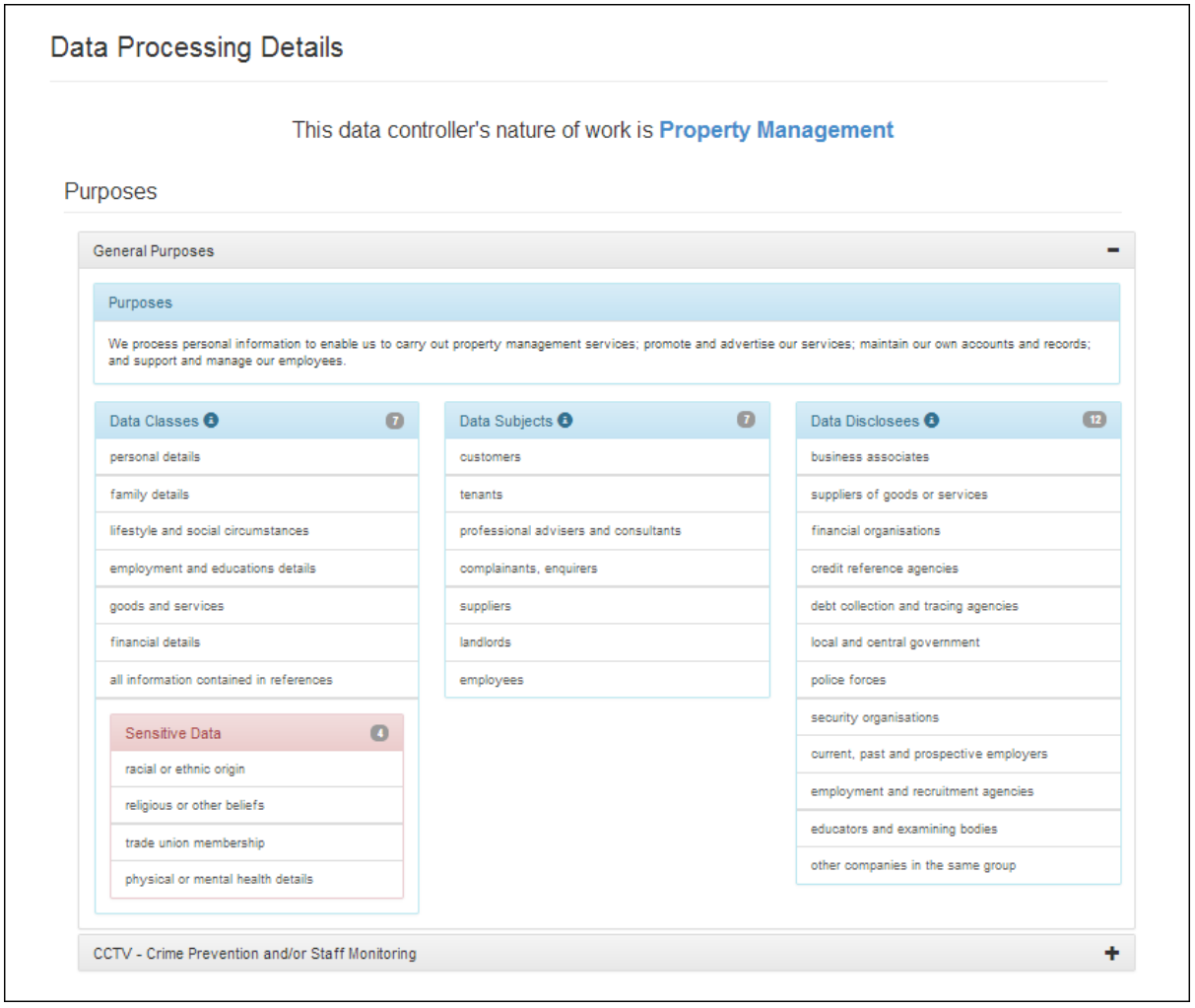

Figure 23. Data processing details

5.4 Third Iteration

We had been successful in making a neat interface for our solution. We now needed to add further richness to our data controller pages and allow for a connected flow of information instead of static blocks.

5.4.1 Statistics

We needed a good way to visualise the information stored by our statistic classes. We believed that a user would like to assess a data controller by the information and the amount collected. They would want to know how the amount of information collected by data controller compares to the general average. For the new format, we could compare this information with the average for the data controller's nature of work and for the older one we could compare with the general average for that particular purpose for data collection. Using these values, we could easily construct bar chart, giving the user an overview of the data collected. Another aspect could be the popularity of a data processing detail. This could be the popularity of a data class, data purpose etc., allowing the user to note that a data class requested is uncommon. For this statistic, we decided to show a donut chart, comparing the percentages of data controllers collecting and not collecting a data information item.

We wanted these charts to pop up after the user clicked on certain data items. Because of this, we made all the panel list groups clickable. The panel group headings were also clickable and

would show the overall comparison of the amount of data collected by the data controller. Once these were clicked on, the graph would appear above that panel column, allowing the user view statistics for items on each individual panel column simultaneously. We had a statistics panel right above the panel groups, indicating the user to click on the items below to view different graphs. Once clicked, JavaScript code would run, retrieving the values hidden on the web page and using them to provide pretty data visualisation in the statistics panel.

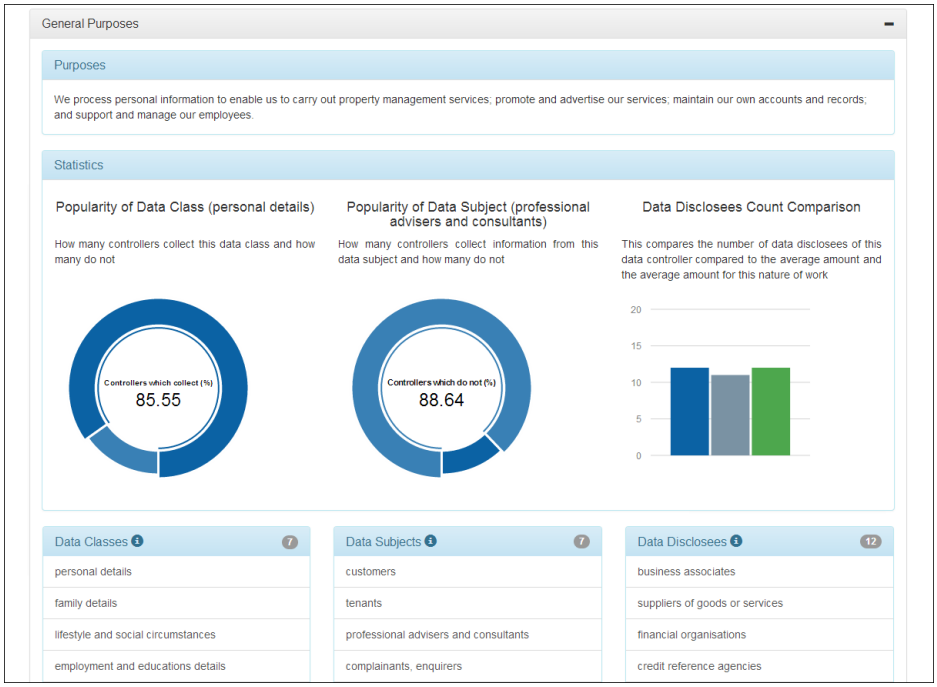

Figure 24. Data visualisations

5.4.2 Linking

We wanted to have more richness in our data. That is one reason we added a Google Map to our data controller page. Another thing we could easily add was information from other resources. If a data controller had a Companies House number, we could retrieve information from other resources. We could work with the Companies House website API and display the extra information that they have on the data controller. Another useful resource was OpenCorporates, which also had more information available if provided with the Companies House number. Unfortunately, there was not enough time to do anything more than providing links to the pages.

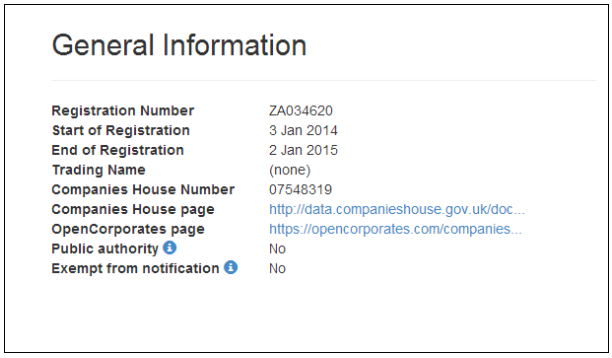

Figure 25. External links in general information group

Another requirement of our project was to link to relevant data controllers from an individual data controller's page. This meant having links to data controllers sharing our current data controller's details. We had this information available to us in our database but we needed a way to effectively use this without cluttering the users display. With this in mind, we thought of having the links to similar data controllers for each individual details with the statistics. This meant that whenever the user clicked on a data class item, they would be shown the popularity of that item along with the link to view all the data controllers which collect that data item.

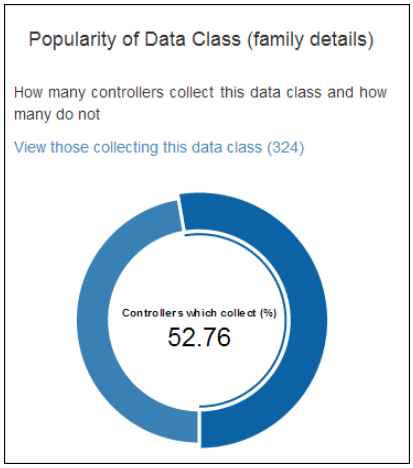

Figure 26. Link to similar data controllers

Once the user clicked on the link, they would be redirected to the list of the data controllers collecting that data item. This allowed us to provide a link for similar companies for each different data item possible, thus giving us a great number of connections with many different companies from a single data controller's page.

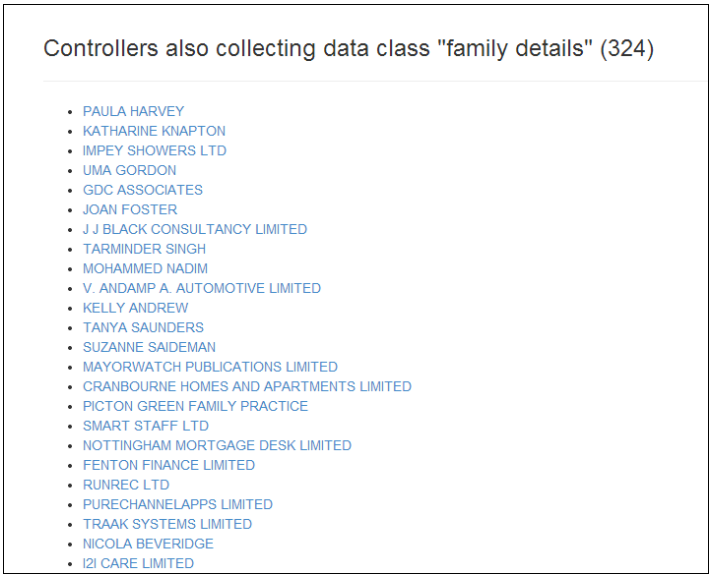

Figure 27. List of similar data controllers

6 Testing

To make sure that different parts of our project work properly, we designed tests to make sure that they were taking the expected actions. However, as we were given the data files by the registry, it was difficult to know what exactly we could do to test our project properly. In the end, the most obvious things to check were how our system would handle badly formed data.

6.1 Methodology

Initially, most of the testing was carried out on our parser. This was because our website just presents the data it has access to. The information retrieved is present in the database and our parser is responsible for filling the database with this information.

6.1.1 White Box Testing

While we improved our parser in the second iterations, we employed white box testing. This was done in order to find out the weaknesses and errors in our methods. Using the Eclipse debug mode, we stepped through our methods to make sure the correct path was taken for each string component while parsing the data processor details. We also employed white box testing on our error cases; when a data controller record caused an error in our parser, we to used white box testing to follow our program's path, identify the problem with its logic and fix it for better accuracy.

6.1.2 Black Box Testing

We also conducted black box testing on our parser program. This was done with a large test script running over the whole registry. If the parser ran through the data processor details without throwing an error, that data controller was considered successfully parsed. Otherwise, the data processing details of that record were written into an error file. This file was later studied and each case was run through the process of white box testing and logic correction where possible.

6.1.3 Unit Testing

Our test was high-level and did not cover the different ways our data could be in a different format than ideally expected. While this would pass through parser, there may remain inconsistencies. Therefore, we decided to test all the different ways data could be different from the ideal format. These cases came from the different errors we encountered initially during our black box testing and further study of different data controller records present in our register. We aimed to make our testing as exhaustive as possible, coming up with a number of ways data could be present in our register file. These were run in an iterative manner, thereby allowing us to correct unexpected behaviour in case of a test case failure.

The different test cases and their results are available in the Appendix.

6.2 Test Outcomes

We ran the large test script to see how many data controllers would throw an error with our parser. With the most robust version of our parser, only 30 records threw an error out of 380,000 data controllers, nearly a 100% accuracy. These 30 had been left after careful white box testing and corrections because they lacked vital information which would be needed to have a proper data controller page. However, as mentioned in our test cases, these cases would be stored in plain html and displayed as is.

7 Evaluation

We needed to find out if our project is better than the currently existing product (the ICO website). Consequently, we ran a user evaluation and used responses to gauge the success of our project.

7.1 Aims

We wanted to understand a few things from our evaluation. We obviously wanted to find out if our representation of data controllers was better than the ICO website. We divided this into a number of things.

- Ease of navigation
- Usefulness of data controllers statistics
- Visual appeal of data controller pages

We also wanted to know the different ways users would want to use a website containing data controller details and if they would use it regularly. A positive response to this would give us more reason to work diligently to provide this useable resource for the general public.

7.2 Methodology

We decided to carry out our evaluations and answer our high level questions by having participants complete a task-based activity and answer a questionnaire. These tasks would be in the form of questions, asking the participants to find out details. They would have to experiment with the different features of the website to reach their goal, filling out the questionnaire at the end of the activity. They would write about the difficulties they faced in completing their tasks, giving us a good idea of how user-friendly and intuitive our features are. They would also give valuable feedback about the different features of the website and offer useful suggestions for improvement.

7.2.1 Tasks

We wanted to perform a comparative analysis of our website and the already existing ICO website as we wanted to know if our created solution was better. In the tasks, we asked the participants to visit two different data controllers and find out varying details about them. With this exercise, we hoped the participants would explore each website to find the required information, forming a valuable opinion in the process. This also served to highlight the differences in the representation of information between the two websites, and also how tedious it was to access it. While we tried to keep the tasks for each website similar, we had to include extra tasks for our website which were related to finding similar data controllers and viewing data controller statistics. This was not possible in the ICO website and they were essential in evaluating our project. To prevent bias towards our website, we randomised the orders of the website the each participant visited first.

The full set of tasks are available in the Appendix.

7.2.2 Questionnaire

Our questionnaire was split into two parts, one for each website. We divided the questions into a mix of qualitative and quantitative ones. The quantitative questions consisted of the asking the user to choose how difficult each website was to navigate, how visually appealing was it, what would they rate it. For the PrivacyMatters website, we also asked the users to choose how useful the statistics and linked data controllers were and how likely they were to use such a

website in everyday life. These allowed us to objectively conclude if our website was easier to navigate through, made for a better viewing and provided a useful perspective of each data controller. It also allowed us to find out if this website could be a valuable resource for the public.

The qualitative questions would allow the users to be more specific about their experience. While questions with objective answers tell us about their preferences, these questions allowed us to get personal opinions. The users were asked to tell us what they liked about each website, what they disliked and to offer suggestions for improvement. They were also asked to tell us how they might use our website generally. This would help pinpoint specific features which can be considered a great success and those are found lacking. We can also find out the different useful features which can be added while the last question allows us to better understand the benefit people would get out of our website.

The full set of the questions present in the questionnaire are available in the Appendix.

7.3 Results

We targeted university students and colleagues who were approached in an informal way. They were made aware of all the different things they would have to do in this study and signed a consent form. They were then linked to the survey which they could carry out at their own leisure. Overll, 17 people answered our questionnaire.

7.3.1 Quantitative Questions

For the ICO website, 47% of the participants found it moderately difficult to find the purposes for collecting data. The percentage of participants who found it less difficult to locate the purposes than this majority was 23% while those who found it more difficult was 30%. Navigating through this website was found to be difficult, with 53% of the participants finding it more difficult than normal while no participant found it very easy. A similar pattern was seen in the question about the tediousness to navigate through the purposes in the tasks for Arsenal football club, with the vast majority of 64% finding it more than moderately tedious. When asked to choose the visual appeal of the website, the response was negative, with 47% participants finding it not appealing at all and another 41% find it less than moderately appealing. Two participants found it more appealing than normal, with one of them finding it very appealing. When asked to rate the website, 60% gave it a 2/5 rating, 22% gave it a 3/5, 12% a 4/5 and the remaining 6% a 5/5.

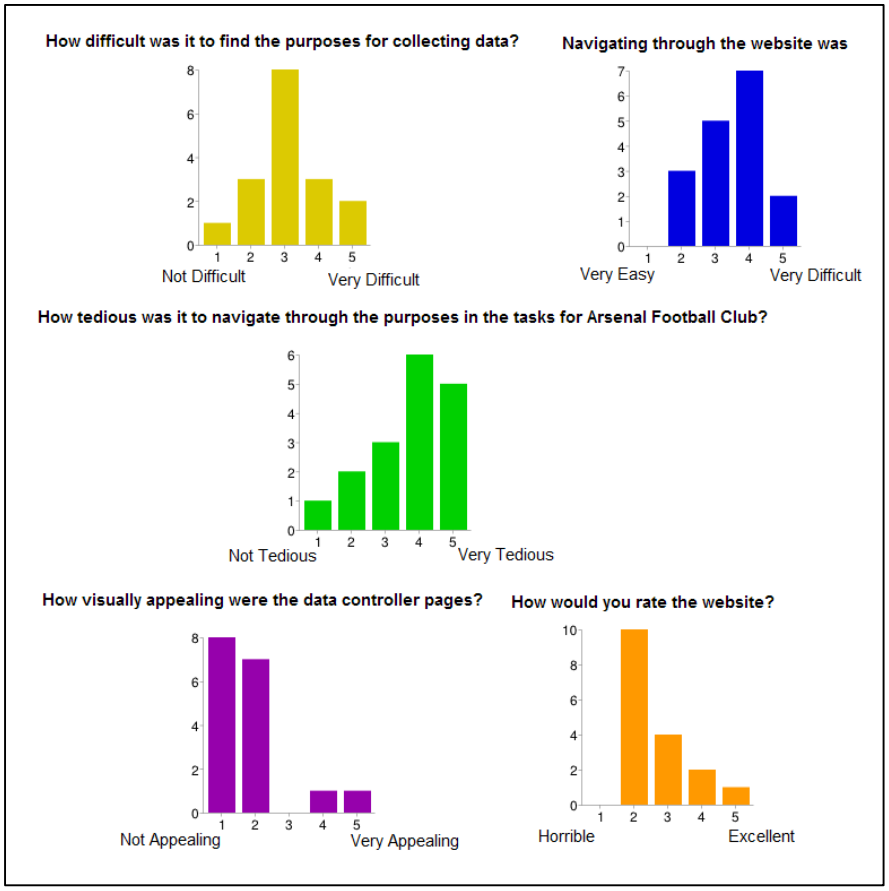

Figure 28. Quantitative results for ICO Website

The results for the PrivacyMatters were contrasting. 35% of participants found navigating through the purposes in the tasks for Arsenal Football Club not tedious at all while an overall 82% found it less tedious than normal. 53% of participants found the website very visually appealing while only 6% of participants found it less than moderately appealing; they found it not appealing at all. With regard to ease of navigation, 77% of participants found it easier to navigate than normal while 12% found it more difficult than normal. According to the majority of the participants (83%), the statistics for each data controller were found more than moderately useful and no one thought that finding similar data controllers in the tasks was less than moderately easy. Those who thought it easier than normal were 89%. When asked about the likelihood of using the website regularly, only 6% thought they were very likely to use it while the rest of the participants were equally divided on the 4 lesser options. Overall, no one gave the website a rating less than 3/5, with 6% giving it that, 65% giving it 4/5 and 29% giving it a 5/5 rating.

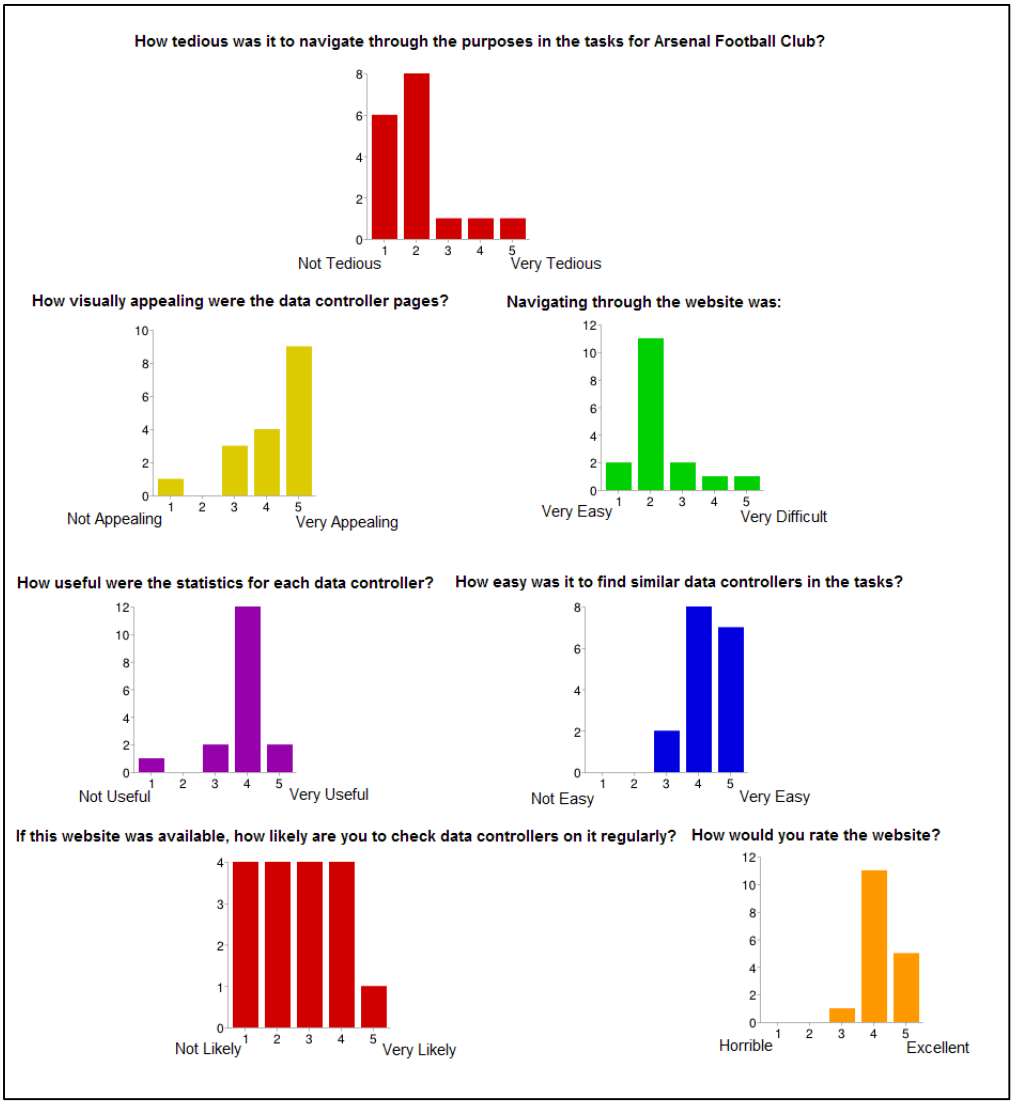

Figure 29. Quantitative results on PrivacyMatters website

7.3.2 Qualitative answers

For the ICO website, people found it simple to use and loved pages loading quickly. They struggled to find anything else with this question answered with an average length of 8 words per participant. While highlighting dislikes, the average answer length more than double to 16 words. In this, participants generally found the lack of any structure boring and unaesthetic. They believed the data was difficult to navigate through, which was further increased by the lack of navigational buttons, wasting time in finding information. For improvements, better navigation, use of a neater layout and more visually appealing interface was suggested. The average length of answers for this section was 15 words per answer.

The participants generally liked the well-structured layout of the PrivacyMatters website, making it easy to find information. Navigation was also performed easily and a few people praised the statistics. This section had 21 words on average. When it came to talk about dislikes, the vast majority had a problem with the 'Chart will be displayed here' placeholder, initially thinking that the section was malfunctioning. Some found was ambiguity regarding where they had to click to show the different statistics and others were not happy that clicking on one purpose resulted in other purpose panel closing. The average amount of words was the same

as the previous section but both were less compared to the improvements section, which had 30 words per answer. The main suggestion was making website more user-friendliness, adding graph icons to imply showing of statistics and removing the 'Chart will be displayed here' placeholder, instead pre-loading charts. When asked about additional information they wanted to see, participants generally felt that the information given was sufficient but a few requested financial data on the data controllers and occurrences of mishandled data. This section had the lowest average words, being 12 words per participant. The last section, regarding ways to use it in everyday life, participants wrote 20 words on average, but many of them did not think they were likely to use it. Those who did, generally wanted to look up information on a data controller they interacted with or were going to. Two participants, however, wanted to use the website to look up different companies to invest in.

The results obtained from the questionnaire are available in the Appendix.

7.4 Analysis

Using our results, we can draw conclusions about our aims and see how well we have answered our questions. It must be said that we cannot draw generic conclusions due to the small number and specific type of participants.

7.4.1 Ease of Navigation

Considering the quantitative feedback of the participants, we can safely say that our website is much easier to navigate through than the ICO website as the average score for PrivacyMatters was less than the ICO website on the difficulty scale. Though it is found 'very straightforward to use' and 'works quickly', 'the lack of navigation buttons' was disliked. In comparison, the PrivacyMatters website is found 'quite easy to navigate through different purposes' and even generally, 'much easier to navigate through'.

7.4.2 Usefulness of data controller statistics

Since the 83% of participants found the statistics more useful than normal, achieving an average score of 3.8 in usefulness, we believe that the statistics were considered to be useful. A few users also mentioned this feature as something they liked finding the statistics 'effective in providing a visual overview of the data'.

7.4.3 Visual appeal of data controller pages

As mentioned before, the simple interface of the ICO website was often praised. However, most of the people had an issue with the lack of structure, as they felt 'Data is presented in a single list, hard to read'. It was also found 'Boring, difficult to distinguish between sections' and that 'Aggregate information is not available'. Its visual appeal resulted in an average score of 1.8 while PrivacyMatters was given an average score of 4.2. Many participants also commented on the interface of our website, mentioning that 'it was easier to find information', finding it 'clean, efficient' and felt that it 'looks nice'. We can therefore conclude that our website had greater appeal than the ICO website and in general.

7.4.4 Usability of website as a resource

The feedback received from the participants was mixed. Many of the participants wrote that they 'probably wouldn't' use this resource in everyday life while the response to the quantitative question was generally negative, achieving an average score of 2.6. Many participants did have uses for this resources, ranging from 'search companies that I use to

find out what information they will collect about me' to 'gain an overview of a company prior to making an investment in it' but it is not something they would regularly do.

7.5 Project Schedule

We created an initial schedule when we submitted our progress report in the form of a Gantt chart. This has been displayed below along with a contrasting diagram of how the actual work was spread out. For many reasons, it cuts a different figure from the Gantt chart that we had initially devised.

The initial 'further planning' block went as expected but the first iteration of the implementation took longer than expected. This was due to the issues we encountered in parsing our data. Three weeks were spent instead of one, setting ourselves behind schedule. Fortunately, the deployment of the website took less time than expected and we got back on track, improving our interface and adding statistics. However, near the end of March, an unexpected personal event came up which required our immediate attendance, lasting two weeks. This caused us to re-evaluate our project and set different goals, essentially to make sure linking and statistics were at the very least functional. This was easily done and some basic features were added before finishing off the implementation for evaluation and report writing. This was another part of our project which had not been properly covered by the by the project schedule before.

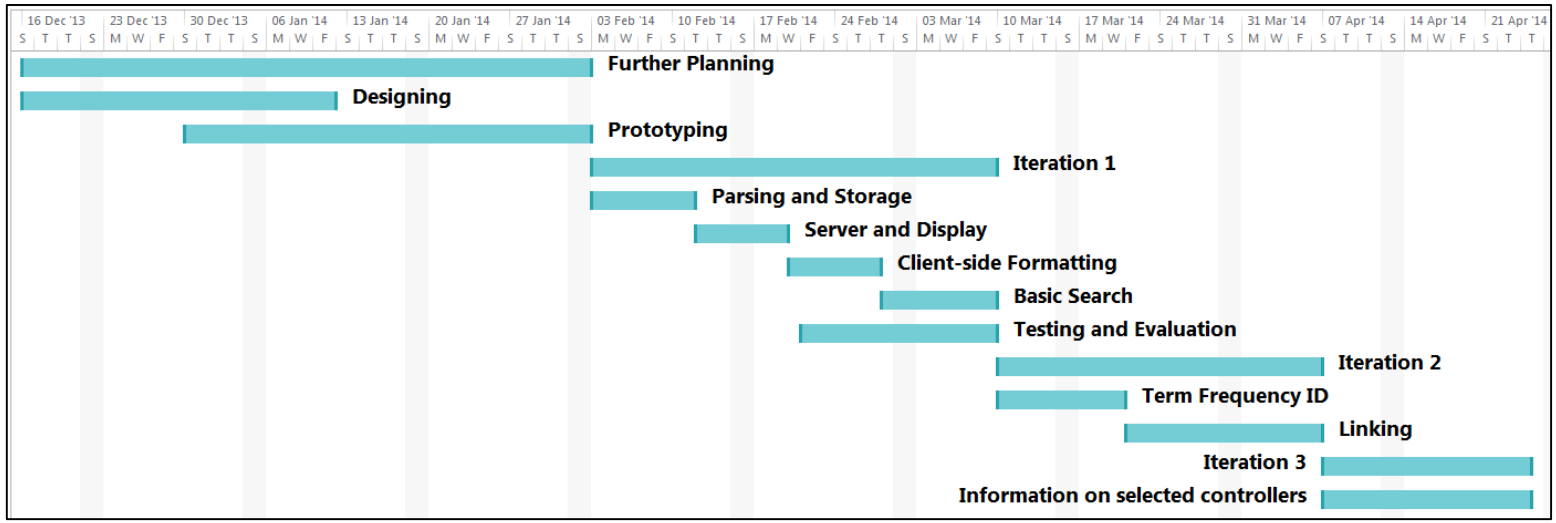

Figure 31. Initial Gantt chart

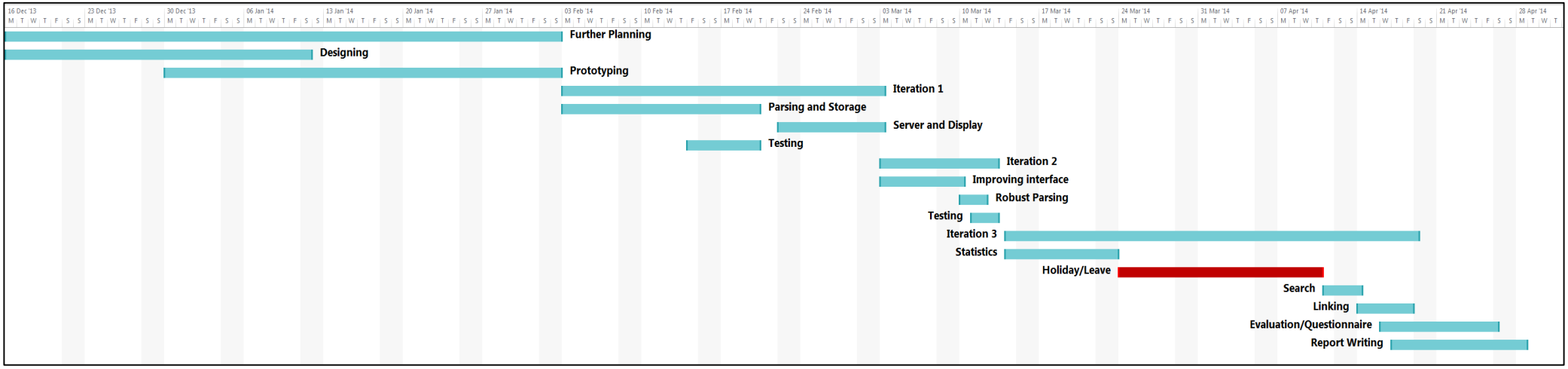

Figure 30. Final Gantt chart

8 Conclusion

8.1 Findings

There were numerous issues encountered with the data register file. The ICO recently changed the format decreasing the richness of our data and denying us a useful way of filtering our data controllers. Instead, we get more distinction in data class items between sensitive data classes and other data classes and have the added Nature of Work attribute which. Ultimately, we would prefer if a mixture of the two formats be achieved; we get the old distinction of purposes but also the nature of work attribute and sensitive data classes.

There is also a need to be objective about the attributes. There are many data controller which list their purposes and details in a prose form. This makes it difficult for a user to better understand the information and our project itself loses the richness of statistics. Be it either format though, it would be preferable if the attributes in their proper tags instead of proving a big HTML block of code and requiring us to parse through it all ourselves. If it were possible for the ICO to have its data available as linked open data, it would be great but that does not look likely.

If we study the results of our user study, we can safely say that we managed to complete our project objectives. Our website offers all the information that the ICO website does, but in a much neater and structured format. Users have preferred our website to the ICO one and greatly admired the statistics and the ability to point to different data controllers from each page. Our system for updating our register is also quite simple as all one needs to do is provide it with the newest register file and let it build the register and the statistics. This would interact with the database and not affect the front-end. With all these things in mind, we can consider our project to be a success.

Nevertheless, there were a few things users disliked and a few suggestions which gave food for thought.

8.2 Expansions and Future Work

This project has added further richness to what already existed with the help of maps and statistics but there is always room for improvement. We will now discuss potential improvements for the future.

8.2.1 Suggested Improvements

Firstly, we should consider the improvements offered by the participants. The most obvious one was making the data controller pages more user-friendly. In its current form, there exist a number of tool-tips and pop-overs to explain main points of the data that is being displayed. However, these could be improved upon. One major change would be icons added to each data list-items, portraying a chart. This would imply the functionality of a graph and we do not need to fill our statistics panel with guidance regarding the interaction with these list items. This would also reduce the confusion faced by the users about clicking on the panel heading to view the comparison of the median information gathered. We could also have a pre-loaded graph in place instead of place-holder text thereby reducing confusion regarding the functionality of our charts and giving a neater outlook of the data controller page. We had intended on having this functionality from the very start but we were unable to get it functioning in time. We could also add more statistics from our current data like gauging the popularity of a data item for a specific purpose or nature of work. A page with overall statistics can be added, showing the

most and least popular data class, subject etc. Lastly, there is also room for us to improve our linking of data controllers. Currently, we only list the data controllers sharing a certain attribute but this can be improved upon by adding further filtering such as nature of work and other attributes too.

8.2.2 Linked Data

One initial goal for our project was to have all our data available as Open Linked Data and thus turn our website into a 5 star data source. Unfortunately, insufficient experience and the priority of our features made this goal extremely difficult to be fulfilled and therefore it was dropped. In the future, we can research this to make it a reality. Currently, our system has functionality for it to be called a 3-star data source; while this has not been made apparent on the website itself, we can display our data controller in a non-propriety (JSON) format. We can further extend this to use URIs to name different attributes and using standards such as RDF and SPARQL. Our website suddenly becomes so much more than just a register to display neat visualisations, allowing other to point to our data and make use of it to create their own visualisations. This way, it may be possible for someone else to make a browser plug-in to show instant information on a company on any website, as requested by one participant.

8.2.3 Further Interactivity

There were a few other things we wanted to implement with our project but they were not possible in the given time. The most wanted feature was the functionality to select a number of companies and show combined information on them. This could allow the user to select all the companies they interact with and be able to view all the information stored on them. They could also view more information such as which company out of all of them collected the most information and many other interesting bits of information. Additionally, if the information we receive is in prose form, we could run analyses on that blocks of text and use intelligent algorithms to extract keywords and useful information out of it. Another useful feature that could be implemented could be a comparison between two or more data controllers, which would be used by a user to better determine which data controller best suits their preferences. One last innovative feature which could be implemented is a grading system. We could set different criteria to what makes a good data controller (say, one which collects the least information) and then give it a rating from A to F. We could grade all our data controllers and provide to our users to filter and view data controllers in another unique way.

8.3 Reflections

Over the course of eight months, we have managed to study one source of valuable data and improve on it. We were able to work with horribly formed data but we managed to extract the useful information out of it and display it in a neater, more understandable manner. We gained valuable experience of working on a big project and learnt valuable skills in web and software while designing our website, which was found was found by many users to be a good user experience while adding further richness and providing a unique perspective to the data. We have also succeeded in laying the foundations of a great resource which can only be improved further, having the potential to become a great utility for the general public.

References

Act, D. P., 1998. *The Data Protection Act 1998.*

Anton, A. I. et al., 2004. Financial privacy policies and the need for standardization.. *Security & Privacy,* 2(2), pp. 36-45.

Beatty, P., Reat, I., Dick, S. & Miller, J., 2007. P3P adoption on e-Commerce web sites: a survey and analysis. *IEEE Internet Computing,* 11(2), pp. 65-71.

Berners-Lee, T., 2006. *Design Issues: Linked Data.* [Online]
Available at: http://www.w3.org/DesignIssues/LinkedData.html
[Accessed 9 December 2013].

Bizer, C., 2009. The emerging web of linked data. *IEEE Intelligent Systems,* 24(5), pp. 87-92.

Bizer, C., Heath, T. & Berners-Lee, T., 2009. Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS),* 5(3), pp. 1-22.

Cattel, R., 2011. Scalable SQL and NoSQL data stores. *ACM SIGMOD Record,* 39(4), pp. 12-27.

Commission, E., 1995. *Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.*

Couchbase, 2013. *Why NoSQL?.*

Cranor, L. F., 2003. P3P: Making privacy policies more useful. *IEEE Security & Privacy,* 1(6), pp. 50-55.

Cranor, L. et al., 2002. The platform for privacy preferences 1.0 (P3P1. 0) specification. *W3C recommendation.*

Fielding, R. T. et al., 1999. Hypertext Transfer Protocol -- HTTP/1.1.

Fischer-Hübner, S., 2001. Platform for Privacy Preferences (P3P) and the Open Profiling Standard (OPS), Draft Opinion of the Working Party: OPINION 1/98. In: *IT-security and privacy: design and use of privacy-enhancing security mechanisms.* Springer-Verlag.

Jacobs, I. & Walsh, N., 2004. Architecture of the world wide web.

Masinter, L., Berners-Lee, T. & Fielding, R. T., 2005. Uniform resource identifier (URI): Generic syntax.

Milne, G. R. & Culnan, M. J., 2004. Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices.. *Journal of Interactive Marketing,* 18(3), pp. 15-29.

Miyazaki, A. D. & Fernandez, A., 2000. Internet privacy and security: An examination of online retailer disclosures. *Journal of Public Policy & Marketing,* pp. 54-61.

Pollach, I., 2007. What's wrong with online privacy policies?. *Communications of the ACM,* 50(9), pp. 103-108.

Schwartz, A., 2009. Looking back at P3P: Lessons for the future.

A Project Brief

Mohammad Ali Khan

Supervisor: Dr David Millard

A.1 Helpful visualisations of EU/UK companies' privacy policies

Companies in the EU are required by law to release information they have collected from clients to the public, also explaining how they intend on using it. This information, available through data control registers, is not in a format which could be easily understandable by the public. Therefore, an attempt is being made to use this data, convert it to a more machine-readable form, organise it properly and use it to create helpful, easy-to-understand visualisations which can be viewed by everyone. Extensions include further analysis of data. The goals of this project are:

- Convert the current data available from the data control registers into machine-readable code
- Have sufficient back end to store the information in a database
- Have working, robust website to display different visualisations based on the data
- Additional third party information on the companies to give a more complete picture

However, there are certain limitations to the project. There is not an intention scrutinise the data sets in great depth or detail; the purpose is more to make them easily available and in an understandable form. This is because the focus of this project is to be more on the web engineering and the problems related to that aspect.

B Testing

These are the different test cases with the expected outcomes and the outcomes. They were ran on the website with respect to how the user would be able to view different information.

Test Case	Expected Outcome	Test Outcome
A data controller record with a valid new format for nature of work description	DataController class created and added to database	As expected
A data controller record with valid old format for nature of work description	DataController class created and added to database	As expected
A data controller record with neither format for nature of work description	DataController class added but html stored in it as-is	As expected
Data list items present in prose form	Only one list item displayed which contains all the prose	As expected
Companies House number is present	Links to open corporate and companies house website	As expected
Companies House number is absent	Companies House number is shown ‘not given’, no link items	As expected
Address and Post code given	Map displayed in ‘Contact’ section	As expected
Address and post code absent	Map not displayed in ‘Contact’ section, Address and Post code labelled ‘not given’	As expected
Valid data controller registration number in address bar	Requested data controller displayed	As expected
Invalid data controller registration number in address bar	User redirected to ‘Data Controller not found’ page	As expected
Search result found in database	List of data controllers displayed	As expected
Search results not found	Message ‘no data controllers found’	As expected
Valid similar data controllers query	List of data controllers displayed	As expected
Invalid similar data controller query	Message ‘no similar data controllers found’	As expected

Table 3. Set of test cases and their results

C PrivacyMatters – Questionnaire

C.1 Privacy Matters

Ethics reference number: ERGO/FoPSE /9396
Investigator : Mohammad Ali Khan

Please read this information carefully before deciding to take part in this research.

C.1.1 What is the research about?

This is a third year project is concerned with data controllers and the data controller registry. A data controller determines the purposes for which and the manner in which any personal data are, or are to be, processed. Data classes are the information it collects, data subjects are the people it collects information from and data disclosees are the people it shares information with. The European Data Protection Directive aims to regulate processing of personal data within the European Union by requiring data controllers to provide their national authority with details on their data processing. We already have a registry available and have attempted to represent it in a richer and more informative manner. We would like to now gauge the success of our attempt.

C.1.2 What will happen to me if I take part?

There will be a task sheet provided to you from which you must carry out a number of tasks. At the end of this, there will be a short questionnaire.

C.1.3 Are there any benefits in my taking part?

The feedback provided by you will be vital to the evaluation of the success of this project. You may also learn about the information different data controllers collect.

C.1.4 Are there any risks involved?

None at all.

C.1.5 What happens if I change my mind?

It is possible to withdraw at any time without any sort of consequences.

C.1.6 Will my participation be confidential?

You are promised unlinked anonymity in your participation.

C.1.7 What happens if something goes wrong?

In any sort of trouble, Mohammad Ali Khan can be contacted at mak1g11@soton.ac.uk

C.1.8 Where can I get more information?

Mohammad Ali Khan can be contacted at mak1g11@soton.ac.uk

C.2 Participant Consent Form

I have read and understood the information sheet and have had the opportunity to ask questions about the study.

I agree to take part in this research project and agree for my data to be used for the purpose of this study.

I understand my participation is voluntary and I may withdraw at any time without my legal rights being affected

C.3 Tasks

Please complete these tasks on the Google Chrome web browser on your desktop computer.

C.3.1 Part 1

1. Visit the url http://ico.org.uk/ESDWebPages/Search

2. Find the data controller "ARSENAL FOOTBALL CLUB"

3. Find the purposes for which controller collects data.

4. Find the information this controller collects for 'Trading / Sharing in Personal Information'.

5. Find the people this controller collects information from for 'Accounts & Records'

6. Now find the data controller "ADIDAS".

7. Find the types of information this data controller collects.

8. Find the people whom this data controller shares information with.

9. Find this data controller's transfer policy on its collected data.

C.3.2 Part 2

1. Visit the URL http://privacymatters.herokuapp.com/

2. Find the data controller "ARSENAL FOOTBALL CLUB"

3. Find the purposes for which controller collects data.

4. Find the information this controller collects for 'Trading/Sharing in Personal Information'.

5. Find the people this controller collects information from for 'Accounts & Records'

6. Find how often Staff Administration is cited as a purpose for collecting data.

7. Find three other data controllers which also collect data for 'Accounts and Records'.

8. Find the number of data controllers that collect 'personal details' from their data subjects.

9. See how the number data subjects for 'Advertising, Marketing and Public Relations' compares to the overall median and the median for this purpose.

10. Now find the data controller "ADIDAS".

11. Find the types of information this data controller collects.

12. Find the people whom this data controller shares information with.

13. Find this data controller's transfer policy on its collected data.

14. Find five other data controllers which collect data from their employees.

15. Find how many data controllers share collected information with the central government. View these data controllers.

16. Find the number of data controllers with the same nature of work as this one.

C.4 Questionnaire

A small questionnaire to help evaluate the third year project

C.4.1 ICO Website

This sections is concerned with the ICO website.

How difficult was it to find the purposes for collecting data? (Not difficult at all 1 – 5 Very Difficult)

How tedious was it to navigate through the purposes in the tasks for Arsenal Football Club? (Not tedious at all 1 – 5 Very tedious)

How visually appealing were the data controller pages? (Not appealing at all 1 – 5 Very appealing)

Navigating through the website was (Very easy 1 – 5 Very difficult)

What did you like about the website?

What did you dislike about the website?

What improvements would you suggest?

How would you rate the website? (Horrible 1 – Excellent 5)

C.4.2 PrivacyMatters Website

This sections is concerned with the Privacy Matters website.

How tedious was it to navigate through the purposes in the tasks for Arsenal Football Club? (Not tedious at all 1 – 5 Very tedious)

How visually appealing were the data controller pages? (Not appealing at all 1 – 5 Very appealing)

Navigating through the website was (Very easy 1 – 5 Very difficult)

How useful were the statistics for each data controller? (Not useful at all 1 – 5 Very useful)

How easy was it to find similar data controllers in the tasks? (Not easy at all 1 – 5 Very easy)

What did you like about the website?

What did you dislike about the website?

What improvements would you suggest?

What additional information would you like to see about the data controllers?

How would you use this website in everyday life?

How would you rate the website? (Horrible 1 – Excellent 5)

If this website was available, how likely are you to check data controllers on it regularly? (Not at all likely – 5 Very likely)

D Questionnaire Results

D.1 ICO Website

How difficult was it to find the purposes for collecting data?				
Not difficult		Moderate		Very difficult
1	3	8	3	2

How tedious was it to navigate through the purposes in the tasks for Arsenal Football Club?				
Not tedious		Moderate		Very tedious
1	2	3	6	5

How visually appealing were the data controller pages?				
Not appealing		Moderate		Very appealing
8	7	0	1	1

Navigating through the website was				
Very easy		Moderate		Very difficult
0	3	5	7	2

How would you rate the website?				
Horrible		Moderate		Excellent
0	10	4	2	1

Likes	Dislikes	Improvements
After search it directs you directly to a particular company if it finds a perfect match.	The user interface - old and not very intuitive; Purposes listed in such a way that requires a lot of scrolling - hard to differentiate and find the information you're looking for.	How many people have information collected by a particular data controller?
Simple layout, fairly easy to navigate.	Aggregate information is not available. Can be confusing at times.	Have an aggregate information page.
It is convenient that all the information is gathered in one area.	The lack of navigation buttons, formatting, and the cluttering of the information made it significantly hard to find the information I wanted in good time.	1) Add navigation buttons that link me directly to the purpose section I want. 2) Using basic HCI, format the page in a way that finding the information is intuitive and pleasant. 3) Add some statistics about the data.

Works quickly, data is all displayed at once.	Data is presented in a single list, hard to read.	Needs a better layout for faster navigation.
The searched data was processed quite quickly as the interface is quite simple	It is quite difficult to navigate through the search results to find a particular purpose of the data controller	Include a drop down menu for selecting a purpose, instead of populating the whole page initially. Include a search bar within the data controller, to make it easy to navigate or search for a particular purpose. Any new search can only be made in home page - a search bar should be included in every page.
It has a simple design and all the data was clearly visible	Time is wasted in finding the information that you want	Have more links so that it is easier to navigate
Lots of information available upon request - also quick.	Very 'industrial' - not very visually appealing. Also quite difficult to find information due to the blocks of text on the page.	Make more visually appealing, make it easier to navigate through information.
well structured information	too much information to read	the website should be more appealing; it looks quite boring
Loaded fast, could easily Ctrl+F for parts I was looking for	No cross-referencing between data controllers, page structure and formatting makes it impossible to know what section text falls under.	A much more obvious tree structure, plus better descriptions of each data item (e.g. what does "personal details" actually mean)
Simple	Did not display relevant statistics, unattractive site. Long pages had to be scrolled through to find single piece of information.	Further links required to areas of the page. Similar to a wikipedia page contents.
Succinct, formal display of information	Boring, difficult to distinguish between sections, unaesthetic	More distinguished sections, ability to click to navigate site
It was fast.	It was simply a page with black text and simply listed things which made going between purposes of the data controller difficult.	A list of the purposes of a data controller which is always on the side to allow easier navigation between them.
Relatively large and clear headings.	Difficult to ascertain what information was contained under each heading, made searching for information difficult. Much of the	Descriptions of what is contained in each heading would make it easier to navigate.

	information was not collated and ready for display i.e. "Chart will be displayed here".	
Very straight forward to use (1 step per page)	1) A lot of content on each page. 2) Organisation of content can be improved	Organisation of content
Nothing.	Everything.	Make it like the second website.
Structured layout, information accessible relatively easily.	Very bland layout, would become tedious after repeated use. Figures are an eyesore.	More organisation, through a better structure for selecting categories.
You can search for data controllers. There aren't any other features.	There is no structure to the data displayed, just plain text document, which you have to browse through manually - horribly discouraging.	Having a structured database which people can query and receive more specific results

D.2 PrivacyMatters Website

How tedious was it to navigate through the purposes in the tasks for Arsenal Football Club?				
Not tedious		Moderate		Very tedious
6	8	1	1	1

How visually appealing were the data controller pages?				
Not appealing		Moderate		Very appealing
1	0	3	4	5

Navigating through the website was				
Very easy		Moderate		Very difficult
2	11	2	1	1

How useful were the statistics for each data controller?				
Not useful		Moderate		Very Useful
1	0	2	12	2

How easy was it to find similar data controllers in the tasks?				
Not easy		Moderate		Very easy
0	0	2	8	7

How would you rate the website?				
Horrible		Moderate		Excellent
0	0	1	11	5

If this website was available, how likely are you to check data controllers on it regularly?				
Not likely		Moderate		Very likely
4	4	4	4	1

Likes	Dislikes	Improvements	Additional Information	Use in everyday life
Very simple and in the same time friendly interface. The controllers have a map, which makes it easy to actually visualize where the company is. The purposes are displayed in collapsable panels, which makes them easy to navigate and browse through.	Finding the other data controllers data have a particular purpose was nested under the "more information" link, which did not make too much sense while looking for it.	Improving the website design, make it more intuitive and user friendly.	Some introduction with regards to what Data Controllers and what I can find from looking into purposes would be useful (some people might not know in advance what a data controller is).	I would look for how companies use my personal information and how much data they collect from me.
Everything in one place, and easy to access.	Some bits of information only appear after you perform certain actions. The existence of these bits of information is not made apparent beforehand.	A bit more user friendliness: a quick search bar at the top right, some tooltips when you hover your mouse over things. Justified text.	None	I wouldn't.
I really liked the intuitive UI, the ease in finding information and all the small features that make this website pleasant to use.	I would have preferred a different text font (maybe larger letters or different font).	1) Different font to match the intuitiveness of the rest of the website. 2) More and more varied statistics would increase the	The information provided on the website is adequate. In the future, I would like to see a graphical representation of the flow of information, although this is not crucial.	I am often interested in knowing what information is kept about me whenever I register in a company. Finding that out, however, is usually a very tedious

		website's popularity.		task. I am glad to know there is an intuitive website that will provide this information for me.
The search mechanism is much better, as is the general display of information.	Sometimes it can be hard to know which of the expandable sections to click on to get the correct information.	It would be nice if some chart was displayed from the start, rather than a placeholder (which made me think that the charts were broken). Ideally, information popups should fade when another part of the page is clicked.	None, it seems pretty comprehensive.	I'd use it to check whether a company I was signing up for an account with would treat my data with care.
Data for the data controllers is arranged in a very accessible manner. It is quite easy to navigate through different purposes. Charts/Stats are quite effective in providing a visual overview of the data - which is always useful. Finding similar data controllers or information regarding data subjects/disclosee	A bit of information should be provided regarding accessing further data on: Data classes, subjects and disclosees - that is clicking any one of them in the menu would provide further statistics and information - though it didn't take	In the charts I would prefer % of those who do collect data as active rather than the opposite.	I have no idea - I believe most of the information is present.	As a consumer, probably to see what information will be recorded by the data controller and to whom the data will be disclosed to.

s is very simple and effective.	very long to realise this.			
Looks nice and has a very simplified interface	It is not clear that you have to click on the 'i' icon to bring up the graphs. It is also not clear what each bar on the graph means.	Make the graphs show up immediately,	The company's website	Search companies that I use to find out what information they will collect about me
Much easier to navigate through and very user-friendly.	n/a	Could perhaps be a little bit more colourful. Also be sure to make the grammar absolutely correct on each word, sentence, etc. Otherwise very good.	Financial statements/balance sheet information - or a summary of the stock price of each controller.	To quickly compile information together. Perhaps do gain an overview of a company prior to making an investment in it.
it was a bit easier to find information; well structered	.	maybe a bit more creativity in developing the website	.	I personally wouldn't use it because the subject has no interest to me; but the website is well structured and organised and I think it's a very well map to use.
Very clear structure, finding things is simple. Lots of linking between pages makes things easy. Extra information is provided with some of the fields. All swish and stuff.	CHART WILL BE DISPLAYED HERE Couldn't find any information about mean/median regarding data subjects.	More descriptions of the various data items (this website also doesn't tell me what "personal details" are). Making the statistics box	More statistics, such as an automatic assessment/ranking of how invasive a company is to customers/employees (perhaps just by adding up the data classes and disclosees)	When providing information to a company it would be useful to know how they intend to use it. I would DEFINITELY use this all

	Clicking entries makes the three columns in the statistics box above change. I'm not sure if the three columns in the statistics box directly relate to the three columns beneath.	make more sense (maybe it should be called "details"?). I sort of want a browser plugin that will tell me "by the way this company will sell your email address to spammers" when I sign up for things.		the time if it could automatically warn me of companies that are likely to share my data.
The data is presented in a clean and appealing manner. It is much easier to navigate through the basic information.	It information about how to access statistics is too verbose. In some cases, the format of graph used does not best serve the purpose (Particularly the circular graph used for popularity).	Small icons to click in certain areas that reveal the graphs. E.g. an icon next to the data class "personal details"	I have no experience in this area.	If I was required to enter information in an online form, I would be able to see what the company was doing with my data and who it was being shared with.
Interactivity, designed to be aesthetic, highlighted sections you clicked on when redirecting to a different part of the page	Not obvious how to display the charts	Maybe display the charts from the start	None come to mind	Check who trades in personal information
Was much more visually appealing, the accordion effect of the purposes made it easy to see what they were and	The list of subjects, classes and disclosees were clickable, but the list simply	On the list of subjects, classes and disclosees, add a graph icon that looks like a	N/A	If I needed to know the details a company held, but most likely not.

gave easy access to detail of each. The data was more compact than ICO as the subjects, classes and disclosees are listed in columns alongside each other.	looked like a table and didn't afford clicking until you hover over and the mouse changes and the cells are highlighted. The opening of one purpose closed another. The graphs shown are to do with all the data controllers, but shown on a single data controller page, can be confusing.	button so users know that the item is clickable. Allow multiple sections to be open at a time. Possibly have a separate page for showing statistics which apply to the whole dataset, as it's not really specific to the given controller whose page you are currently on. No details on the Adidas page for Undertaking Research, some other differences between it and the ICO page.		
All information on display straight away without the need to open up the different tabs like in the previous website.	Large blocks of text with nothing to break it up.	Links at the top of the page to jump to each controller along with an index of each purpose. Maybe use different coloured text for each purpose as well to make it easier to	Lack of statistical information on display.	I probably wouldn't.

		differentiate each section at a glance.		
The layout of the UI is neat. The website is fast in data retrieval for the charts. Clear organisation of most content.	1) Search could be improved with auto-suggestions. 2) I had to scroll through the page twice to get a better understandin g of how the charts worked.	I thought that the text 'Charts will be displayed here' was a little misleading. While scrolling through the page I first thought that they were yet to be added to the site. Perhaps you could display the bar graph as default.	-	Probably to know which controllers have access to my data and how they manage it.
Everything.	Nothing.	make the homepage more colourful.	I don't know what data controllers are.	I personally wouldn't use it. However I think businesses would find the information very useful.
Clean, efficient. No need to open all subcategories when viewing data. More chart data, different pictoral representations aid user accessibility.	Categories under purposes could be spaced out a bit more, making it easier to read.	Space out category list, add more colour.	Nothing.	As a search tool to help gather information on companies. Perhaps before choosing to invest.
I liked the statistics and the pages which show how controllers are related with respect to a certain subject, class etc.	Nothing of major note, just some suggestions.	1. Have auto complete in the search on the homepage 2. Show some statistics and engaging graphs on the homepage 3. When viewing	Whether they have misused data in some way, are there any privacy concerns with this company? Do they have a track record of making sure they follow the rules?	I would check if before filling out important personal information with someone what a company does with it.

		information about the data controller have the ability to filter not only by 'Purposes' but by 'Data subject', 'Data classes' and 'Data Disclose'. Say I'm a client I would like to see for what purposes someone collects data from me and what exactly collect. I would not go through all the purposes and look if client is in the data subjects list.		