Samuel Grogan* and Aleecia M. McDonald

Access Denied! Contrasting Data Access in the United States and Ireland

Abstract: The ability of an Internet user to access data collected about himself as a result of his online activity is a key privacy safeguard. Online, data access has been overshadowed by other protections such as notice and choice. This paper describes attitudes about data access. 873 US and Irish Internet users participated in a survey designed to examine views on data access to information held by online companies and data brokers. We observed low levels of awareness of access mechanisms along with a high desire for access in both participant groups. We tested three proposed access systems in keeping with industry programs and regulatory proposals. User response was positive. We conclude that access remains an important privacy protection that is inadequately manifested in practice. Our study provides insight for lawmakers and policymakers, as well as computer scientists who implement these systems.

Keywords: Internet Privacy, Data Access, Privacy, United States Law, Irish Law

DOI 10.1515/popets-2016-0023
Received 2015-11-30; revised 2016-03-01; accepted 2016-03-02.

1 Introduction

The privacy principles outlined in a 1973 HEW report became the foundation of several major privacy approaches [74]. The Organization for Economic Cooperation and Development (OECD) forms the basis for data privacy principles in Europe, based on principles taken from the HEW report [59]. Ireland is a signatory to the OECD guidelines on the Protection of Privacy and Transborder Flows of Personal Data [57]. The five Fair Information Practice Principles (FIPPs) are also based on the HEW report, and inform the United States' approach [30, 44]. These five principles are:

1. Notice, provided by privacy policies;
2. Choice, which imagines users electing to avoid websites with insufficient privacy protections;
3. Access, which empowers users to see what data is held about them;
4. Integrity, addressed by data breach laws, encryption, and security tools;
5. Enforcement, often provided by the FTC itself.

Access is a core privacy protection safeguard, providing consumers with the ability to view the data collected about them. One can see how access rights are written into the US' 1970 Fair Credit Reporting Act, which the Federal Trade Commission (FTC) summarizes in part as "You have the right to know what is in your file" [31]. Many types of data collection, use and disclosure are permissible under the FIPPs if individuals have the ability to self-manage their privacy [65]. By self-management, we mean if they are notified of data collection and use, and they provide consent. Similarly, in Ireland, data collection is permissible and considered to be processed fairly if a data subject is notified and consents prior to data collection [57].

Low priority given to access: Access as a privacy protection is often missing or extremely weak online. Despite access rights earlier prominence in US law, since the 1990s the FTC has particularly emphasized the first two FIPPs concepts of notice and choice rather than access. For example, we speak specifically of a "notice and choice approach," but not of an "access approach" to online privacy. In practice, a website might allow a user access to her home address and email address in order to update them, but omit access to what behavioral information they collect about her let alone any inferences they made. Yet if websites provided raw data, for example a web server log line, it would likely make no sense to users and become unusable due to information overload.

An Internet user who is notified of the types of data a company collects may consent to this data collection, which is relevant both under the US notice and choice approach, and the Irish focus on consent for most secondary uses of personally identifiable information. However, Internet users are often not allowed to access the behavioral information collected about them by a company, nor may they be able to see what is done with that data. Furthermore, if the company compiles a pro-

***Corresponding Author: Samuel Grogan:** Queen Mary, University of London, E-mail: s.r.grogan at hss15.qmul.ac.uk
Aleecia M. McDonald: Non-resident Fellow, Stanford Center for Internet and Society, E-mail: aleecia at aleecia.com

file of that customer and makes errors, the consumer has no recourse to correct that information as they could not know that there is an inaccuracy in the first place. Lack of access rights undermines the other privacy approaches of notice, choice, and consent.

Access and Data Brokers: According to a 2012 FTC report on protecting consumer privacy, data brokers are companies that collect information, including personal information about consumers, from a wide variety of sources for the purpose of reselling such information to their customers for various purposes [34]. Consumers are often unaware of the existence of these entities, as well as the purposes for which they collect and use data [33]. A 2013 report by Senator Rockefeller which reviews the data broker industry notes that "data brokers operate behind a veil of secrecy," [72]. The Senate Commerce Committee revealed that in the course of the study leading up to the report, large data brokers such as Acxiom, Experian and Epsilon refused to reveal the specific data sources they were working with [10]. Acxiom executives have stated that the Acxiom database contains information about 500 million active consumers worldwide, with 1,500 data points per person on average [63].

Access principles could improve consumers' privacy with regard to data brokers. In September 2013, Acxiom gave consumers limited access to data it has collected about them [12], yet there are over 200 data brokers that do not offer access [9]. There exists little information about Irish data brokers.

2 Background

In this section, we discuss why comparing the United States and Ireland is important, detail current access laws and proposed reforms, and examine access tools.

2.1 Why the United States and Ireland?

The United States and Ireland, both common law jurisdictions, approach access differently. By comparing the two legal approaches, we can examine which, if any, is more effective.

Major Internet companies and data brokers are often US based and it is important to examine the environments in which they operate. Ireland is important to the study as it continues to be one of the main European operating locations for technology companies and as a result, Irish law governs these companies. Some European member states have stated they view Ireland's approach to data protection as too lax in order to attract US companies [19, 51]. Ireland's data protection Commission denies these claims [14]. In 2015, Facebook was involved in a dispute with Belgium's privacy protection Commission who accused Facebook of tracking European Internet users without their consent [55]. Facebook argued that they are governed by Irish law and any disputes should be resolved through Irish regulators [39]. Nevertheless, in November 2015, a Brussles court ruled that Facebook must stop tracking non-users of the social network [24]. Facebook plans to appeal [61].

Finally, English is the operational language for both the United States and Ireland. By conducting the study in English we are able to avoid confounds due to language. Conducting the study in two different languages would require extraordinary care in translation to ensure accurate analysis.

2.2 Current Laws

Data protection laws differ world-wide. The US does not have a uniform data protection law. There is also no single regulatory authority dedicated to overseeing data protection law in the United States [66]. The FTC is the primary federal privacy regulator regarding consumer protection. The US does not have any one specific method of protecting a citizen's right to access data collected about them by companies, relying on a patchwork of state and sectorial federal laws for credit agencies and data brokers. Privacy legislation tends to be adopted on an ad hoc basis, with legislation arising when certain sectors and circumstances require [37].

Specific US laws that address access rights include the federal Health Insurance Portability and Accountability Act (HIPPA), Children's Online Privacy Protection Act (COPPA) and California's Shine the Light Law. HIPPA gives individuals the right to access personal health information collected about them [1]. COPPA allows parents or legal guardians to obtain access to the personal information collected online from their children [2]. The State of California's Shine the Light Law gives a California resident the right to access any personal information that has been shared about them with third parties, as well as the names of parties with which the information has been shared [17].

In contrast to the US, the 1995 Data Protection Directive 95/46/EC [58] regulates the processing of personal data within the European Union. The Directive is general and not limited to a narrow area of protection as

in the US. The Directive has been transposed into Irish law by the Data Protection (Amendment) Act 2003 [6]. Section 3 of the Data Protection Act allows an individual to find out free of charge, if a company or individual holds information about them, including a description of the information held and the purposes for which it is held. Section 4 adds a right of access. This section establishes a right to obtain a copy of any information held about them for a nominal fee.

2.3 Proposed Reforms

There have been a number of recommended reforms to data privacy laws in recent years. We will consider the US recommendations and then Ireland's proposals.

The California Assembly bill entitled "The Right to Know Act 2013" proposed to update the right of access established by the Shine the Light Law [16]. The bill would require businesses to disclose what personal information they hold based on a customer's request. Companies must also disclose names and contact information of all third parties with which the business has shared that customer's data during the previous 12 months. Ultimately, it would give California residents the right to access their own data and to see the flow of data between one firm and another. The American Civil Liberties Union of Northern California asserts that this act would modernize current privacy law [8]. However, while this bill updates an existing access law for California residents, the problem still remains that companies themselves may not know where the data flows to. For example, if a website hosts advertising, often the ad space is auctioned off in real-time. A visitor to the website could be tracked by dozens of third parties, only to have a different collection of third parties track her when she returns a week later. The hosting website cannot answer which sites have collected data about a specific user because the first party website does not know itself. This poses a fundamental challenge to the California bill, as well as to the Irish notion that a data controller is responsible for all data processors with access to user data via the data controller. The Right to Know bill failed to pass through the California Assembly Judiciary Committee in January 2014 [16].

In June 2013, Commissioner Brill of the FTC suggested a new program called "Reclaim Your Name" [13], which would establish technical controls allowing people to access the information data brokers, specifically, have stored about them, control how data is shared, and correct incorrect information.

More recently, the White House published a draft of its proposed Consumer Privacy Bill of Rights Act of 2015 (CPBR) [20]. The Act applies to any "covered entity" that collects, creates, processes, retains, uses, or discloses personal data, including data brokers. The bill requires individuals be given access to or an accurate representation of their personal data. However, the CPBR contains limitations that could result in consumers being denied access by data brokers. These include measuring the degree of access against the risks associated with the data and the costs incurred by the covered entity in providing access [47]. The CPBR also limits access requests that are frivolous or vexatious.

Mere days after the Obama administration proposed the CPBR, Senator Markey put forward the Data Broker Accountability and Transparency Act 2015 [4]. The proposed legislation would allow consumers to see and correct personal information held about them by data brokers and to opt out of their data being used for marketing purposes. Unlike the CPBR, Markey's Act applies exclusively to data brokers and requires a data broker to maintain an Internet website to allow individuals to review information about them and to express their preferences [47].

As with California's data access law, the major problem with European data protection laws is that they are outdated. Irish law has been slow to catch up with new technologies. Two decades have passed since the introduction of the Data Protection Directive. Globalization, in addition to new technological services, brought many new challenges. The European Commission produced a report in 2012 outlining a need for EU data protection reforms [28]. This report proposed new rules to update legislation to ensure effective protection of the fundamental right to data protection and improve certainty as to the law for companies. Subsequently, the European Commission put forward the General Data Protection Regulation (GDPR) [21] which will unify data protection law within the European Union. This legislation will apply to Ireland, preempting existing national law. While the GDPR updates data protection law in certain areas, little change has been proposed for access rights of data subjects with more attention given to the right to be forgotten [11]. The GDPR updates the Data Protection Directive's provision on access in two relevant ways. Firstly, data subjects are now entitled to find out from data controllers the periods for which their data will be stored. Secondly, it seems that organizations must now respond to an initial subject access request free of charge and may only charge a reasonable fee for further requests for the same data [22]. The Eu-

ropean Commission, Parliament and Council agreed to the GDPR text in December 2015 [22]. It is expected to come into force in 2018.

The proposals put forward by the United States and Europe could advance data access rights. These proposals tie into our study as we seek to investigate what Internet users know about the current laws and what they would like to see change to improve their right to access data held about them. Ideally, legislators and policy makers will be able to use this information to build on the existing proposals and perhaps draft new recommendations to further develop data access.

2.4 Existing Tools

Although few dedicated access tools exist, many companies are releasing their own privacy tools which aim to increase user trust and provide transparency. Facebook introduced a privacy checkup tool in June 2014 [56]. The company expanded the information they show about ads including why a certain ad is shown. Typically, the user is informed that they are in an ad category that the company responsible for the ad is trying to reach. The tool also integrates an ad manager.

In 2015, Google launched their own privacy hub (see Appendix B, Fig. 14 and Fig. 15) which allows users to limit the kind of ads they see. Account holders can also prevent Google logging their activities while using Google's services and can download a copy of their data from Google. The aim of this hub is to make privacy controls easier to find. Previously, the controls were spread across Google's different services [50].

Privacy managers are not new. Yahoo! introduced their privacy center in 2002, which was updated in 2008 [70]. Although this tool prides itself on giving Internet users greater accessibility to manage their privacy settings, as of June 2015, only the US version of the privacy center provides a link to privacy tools [77]. This is omitted from the Irish version [76]. While Irish users can still access these tools through Yahoo's individual services, the idea of a central privacy center is largely defeated by this omission.

Microsoft incorporates privacy management into its safety and security center [53]. Instructions are provided on how to access privacy settings from each individual Microsoft service. To manage ads, users are redirected to choice.microsoft.com [54].

3 Related Work

In this section we discuss previous work on data access. Many scholars have examined user perceptions of data collection practices but there has been little work conducted on users' attitudes to accessing that data once it has been collected.

Previous research conducted by Cranor et. al. indicates that that the most important factor for users in disclosing information about themselves on a website is whether it will be shared with other companies and organizations [23]. A study by Ur et. al. notes that Internet users believe companies collect more information than they generally do collect [71]. Access rights could build user trust by establishing data collection is more reasonable than feared. In June 2011, the European Commission published a Eurobarometer [68] on attitudes on Data Protection and Electronic Identity in the European Union. The study indicated that 70% of Europeans are concerned that their personal data that is held by companies may be used for a purpose other than for which it was collected. The report also noted that more than a quarter of Europeans are prepared to pay to access their personal information, which is one measure of how much citizens value access rights.

In June 2015, the European Commission published another Eurobarometer on Data Protection [69]. While this study addressed the extent to which Internet users feel in control of the information they provide about themselves online, there were no questions about data access included in the study. Similarly, in May 2015, the Pew Research Center published a report [52] which also focused on user control over their data. Access was unaddressed. While access is an important privacy safeguard, it continues to be given short shrift.

Previous research indicates that while Internet users are concerned about their privacy online they are unwilling to change their behavior to achieve it [5, 67]. However, with regard to access the so-called privacy paradox is questionable. In 2011, Max Schrems made an access request under European law for all the data held about him by Facebook. He received a CD containing 1,200 pages of data held about him by the company [73]. Schrems then created a website which provides users with information on how to access their data from Facebook [27]. Soon after the site's launch, Facebook received 40,000 access requests [64]. This suggests that users are willing to change their behavior to access their data. Perhaps in response, Facebook changed

how they process access requests, and no longer provide access to as much information [27].

The data collected by online companies and data brokers is often personally identifiable information. This collection can be done by either first party or third party companies. The FTC defines first parties as sites you visit and third parties as someone other than sites you visit [32]. Increasingly, in international law, personally identifiable information is defined as any unique identifier. The GDPR defines personal data as any information relating to an identified or identifiable natural person. An identifiable person is one who can be identified, directly or indirectly, by reference to an identifier [22]. The California Business and Professions Code defines personally identifiable information as individually identifiable information about an individual consumer collected online by the operator from that individual and maintained by the operator in an accessible form. This includes any identifier that permits the physical or online contacting of a specific individual [15]. The California Attorney General in reading this law defined personally identifiable data as any data linked to a person or persistently linked to a mobile device — data that can identify a person via personal information or a device via a unique identifier. Included are user-entered data, as well as automatically collected data [48]. Personally identifiable information has been established as any unique identifier at federal level also. COPPA's definition of personal information also includes any other identifier that the Commission determines permits the physical or online contacting of a specific individual [2].

Another aspect of our study is how Internet users view data brokers whose main economic aim is the selling of users' information to third parties. Data broker information can be one source of data for behavioral ad categories. Data brokers are not new. For example, Acxiom was founded in 1969 [38] but it is only recently that we began to learn what data brokers are holding about Internet users. Some examples include lists of rape victims, substance abusers and derogatory credit consumers [25]. In September 2013, Acxiom launched the website aboutthedata.com which gives consumers a limited insight into the personal information Acxiom holds about them, their home, vehicle, economic data, shopping and household interests [7]. Acxiom shows broad categories, not data the company holds on consumers for marketing purposes. Senator Rockefeller notes that consumers do not have access to data which Acxiom has applied analytics [72]. For example, a consumer could see data points showing their occupation and that they have children, but if Acxiom inferred from those data points that the consumer is a "working parent," the consumer would not have access to the underlying data elements [72]. Other data brokers do not allow consumers to view or correct any data they hold about them. As Angwin notes [9], of the data brokers that offer an opt-out (almost always without access or correction), even the simple task of opting out does not meet consumer expectations. Opting out does not mean deleting information, only to "suppress" data. Because there is no right of access, people who use data broker opt outs never know if those opt outs work. There is no way for users to see what data brokers continue to share, which undermines trust in the opt out approach.

Little information exists about Irish data brokers. We emailed the data protection Commission of Ireland for information and they directed us to the Irish Brokers Association [36, 46]. However, the Irish Brokers Association is a representative body for Irish insurance brokers, not data brokers [45].

Our work builds on these studies as we seek to examine views on data access to information held by online companies and data brokers.

4 Research Questions

In this paper, we aim to answer the following research questions:

- How aware are Internet users of current data access mechanisms?
- Are Internet users interested in accessing the data held about them by online companies and do they know how to access their data?
- What do Internet users know about data brokers?
- Are Internet users interested in using the systems we propose which would allow them to work with data held about them by companies and data brokers?
- Do attitudes and knowledge of data access differ between American and Irish Internet users?

5 Methods

We conducted a survey to analyze user awareness and attitudes around data access.[1] The survey consisted of 40 questions on average and took under 15 minutes to

1 This work was approved by Stanford University's Institutional Review Board as an exempt study.

complete. The survey also included three possible interfaces to access data categories. All survey responses were anonymous but we asked each participant which country they were located in. This allowed us to compare the US and Irish answers. We ran a pilot study on Amazon's Mechanical Turk (MTurk[2]), a crowd sourcing service that is also a popular tool for use in human-computer interaction research [49]. The pilot study ran from 6 August 2013 until 31 December 2013. The final study ran from 13 April 2015 to 20 April 2015.

5.1 Pilot Study

We presented our survey as a "task" on MTurk. We received 605 valid responses. Based on participant responses and feedback we revised our questions and UI screenshot designs for the final study. We learned that MTurk, while proven to be useful in many studies, is not the appropriate platform to gain representative samples of the general US and Irish populations. We had difficulty in recruiting Irish respondents due to the way MTurk operates, and limitations on selecting participants by region. In addition, most users of the service are US based. In light of this, we chose to use panel services to recruit our participants for the final study. After one year of further revision, we launched the final study in April, 2015; we report those results below.

5.2 Participants

We recruited participants through CINT[3] [18], a SurveyGizmo panel company partner. Our recruitment materials indicated that participants would complete a survey on online data practices. We did not mention privacy to avoid selection bias and participant priming. Each participant who completed the study was paid according to the panel company's market rate.

We took a number of measures to ensure that our study participants were answering the survey seriously. Firstly, we designed the survey so that it is clear from the beginning that the study was not a quick task. Progress bars displayed during each question of the survey and this tool encouraged those not taking the survey seriously to exit. Secondly, we discarded any participant who answered the survey in less than 5 minutes. The majority of participants who completed the survey under this time gave nonsense answers to long text response questions and those that made sense were often inconsistent with answers given to related questions. This practice stopped for most participants who completed the survey in 5 minutes or more. We conducted a further filtration process among the remaining participants to ensure validity of all responses. This involved reading the long text boxes, and comment section for every participant and manually filtering out invalid responses. Often times, participants who took the study seriously left comments at the end of the study with some further thoughts on data access and collection. This greatly aided the filtration process. 1410 participants engaged with in our study. 390 participants (28%) dropped out before completing the survey. This rate was largely due to one rogue participant who attempted to take the study numerous times in order to be compensated multiple times. However, the panel company has systems in place to avoid this and therefore, the rogue participant was forced to drop out after each attempt. After manually checking the remaining 1020 responses and filtering out invalid or too rapid responses, we had 873 valid responses. These consist of 431 valid US survey responses and 442 valid Irish survey responses. Information about demographics is in Appendix A.

5.3 Questions and Analysis

We defined "access" and "data brokers" to participants in the survey and made clear to them what types of data the questions are asking about.

We defined "data access" as the right you have to access the data, or information, held about you by others. Participants were told that data access rights in relation to Internet users specifically refers to the users' right to access their personal data collected by companies as a result of the users' activities. We defined "data" for participants by informing them that any reference to "your data", "data held about you", "data collected about you", "personal information", etc. refers specifically to data collected about you as a result of your online activities. We then provided the following examples of data: Name, Phone Number, IP Address, Email Address, Health Information, Financial Information, Nationality, Work History, Web Pages Visited, Calendar Application. We indicated to participants that the list was not exhaustive and we acknowledge that the types of data collected can go far beyond our examples. We chose these examples because many companies use the

2 http://mturk.amazon.com

3 CINT partnered with Lightspeed GMI in order to source Irish participants.

same examples in their definition of personally identifiable information [41, 77]. These examples are also present in many laws and regulations [2, 15, 22]. We defined a "data broker" as a company that collects information about individuals and then sells this information to other companies for various purposes. For example, marketing purposes. We chose this definition as this is how the FTC defines a data broker [34].

We used a combination of question formats to gain a better understanding of participants' awareness of data access and their desirability to view data held about them by companies. We also sought to explore what Internet users know about data brokers. The survey questions are in Appendix A. Within a given question, the order of suggested answers was randomized for each participant to further ensure the validity of results. We examined whether significant differences existed between the responses of Internet users from each population; please see Appendix C for details.

5.4 Study Limitations

Despite the fact that the responses received were from people of both genders, across many different age groups, and employed in many different industries, our survey participants were all recruited from a panel company and thus our results may not necessarily be representative of the whole US and Irish populations, despite attempts to recruit a representative sample.[4] We are analyzing self-reported data and user responses may not necessarily reflect their actions in practice.

6 Results

In this section we present our results. We discuss implications of these findings below.

Of the 873 responses, 48% of respondents were male and 52% were female. 28% of respondents were aged 18-34, 38% aged 35-54, and 34% of respondents were aged 55 and over. Questions regarding age and gender appeared at the start of the survey; more detailed demographic questions such as education level and income level appeared at the end (See Appendix A).

4 We chose our sample size after consulting CINT who advised us on which sample sizes would likely be representative of the US and Irish populations.

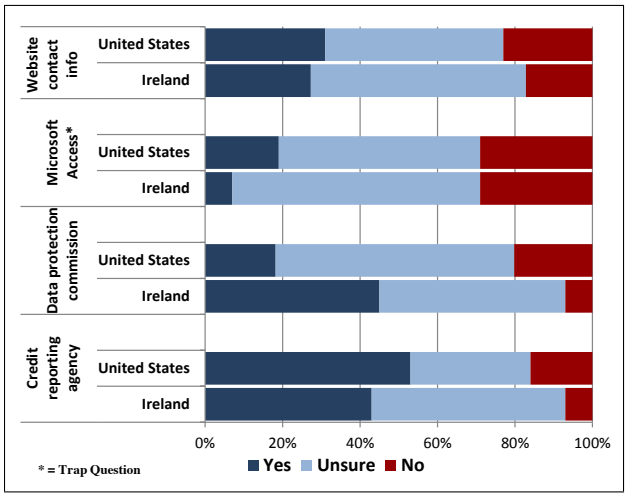

Fig. 1. Awareness of data access through various sources.

6.1 Awareness of Access

In this section, we focus on participants' responses to questions which seek to gain an understanding of their knowledge of data access law.

Sources of Access: We began the survey by asking "Can you access data held about you through the following?" and provided four possible answers.

We found a statistically significant differences between US and Irish participants for every category of this question (p=.025). As seen in Fig. 1, more of our American participants are aware that they can access data about themselves by writing to a credit reporting agency, as established by US law, than our Irish participants who have the same right of access through the Irish Credit Bureau (53% US, 43% Ireland; p<.000). 45% of Irish participants correctly identified that they may access their data by writing to the Data Protection Commission, which is the main entity for Irish citizens to access their data. 18% of American participants mistakenly believe they can access their data this way (p<.000). A few, but not all, websites offer data access and nearly one third of our US and Irish participants believe that they can gain access to data this way (31% US, 27% Ireland; p=.016). The program Microsoft Access is a database package, yet in both regions some participants believe they can use it to access their data (19% US, 7% Ireland) or are unsure (52% US, 65% Ireland; p<.000). We find that US participants are more likely to know about their limited right to access credit information than Irish participants are to know about their expanded right of access through the Data Protection Commission, but in each region people are confused over where to go for data access.

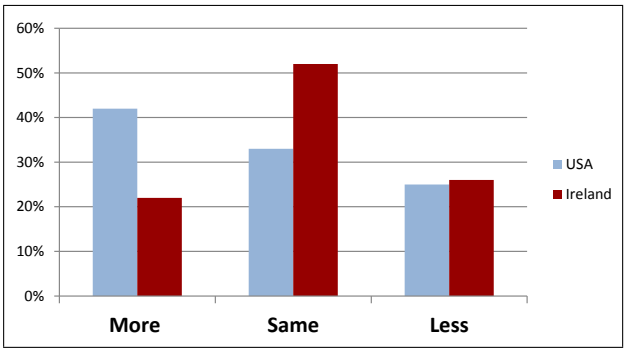

Fig. 2. Participant views on degree of access their laws provide compared to other countries.

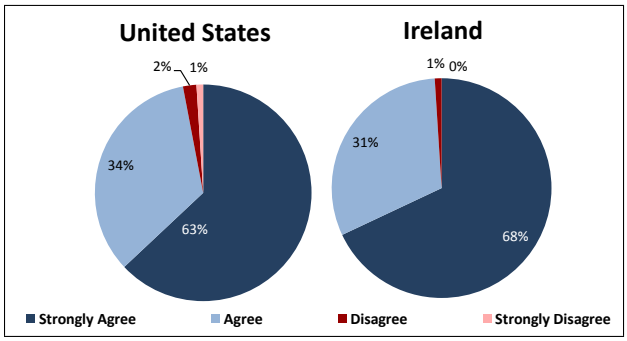

Fig. 3. Level of desirability for the right to access data.

Knowledge of Existing Laws: In certain cases, Irish citizens are able to access their data with less difficulty than Americans thanks to the 1995 EU Data Protection Directive [58]. We asked our participants whether they believe data access laws in their country give them access to more, less or the same types of information held about them by companies in comparison to other countries. Our American participants are more inclined to believe that US laws allow for access to more information held by companies than laws in other countries (42%). Irish participants are significantly less likely to believe that Irish laws allow for access to more information (22%; $p<.000$) as shown in Fig. 2. Overall, we find American participants think they have stronger access rights than other countries, and that Irish participants are as likely to think their rights are merely on par with other countries.

Using a free form text response, we asked participants to name a law which exists to protect their right to access data held about them by companies. 40% of Americans reported that they did not know of any. Following this, the most popular responses were the Freedom of Information Act, which allows US citizens to gain access to governmental rather than company records (8%), the Data Protection Act, which is not a law of the United States (6%) and HIPPA (6%). Several participants also reported that a law protecting their rights to access does not exist (6%). 51% of Irish participants correctly answered with the Data Protection Act which is Ireland's primary access legislation. 26% of Irish participants reported that they did not know of a data access law. The Freedom of Information Act, an Irish law that serves the same purpose as the identically named US law, was the next most popular response (8%). We find that two fifths of American participants and one fifth of Irish participants are unable to identify a law that protects their right of access. Those who can identify an access law appear to think laws which grant access to government records can also be used to access data from companies.

6.2 Interest in Access

In this section, we examine levels of desirability of data access. We consider to what degree participants have tried to access their data and the effort participants are willing to go in order to access held about them.

Desire for Right of Access vs. Desire to Utilize Right of Access: We asked 50% of our participants whether they wanted a right to access data a company holds about them. The other 50% were asked whether they would be interested in utilizing such a right to see the data companies hold about them. This allowed us to compare users' desirability of having *the right* to access their data vs. whether users are interested in actually *utilizing* the right to access data held about them.

As shown in Fig. 3, participants from both populations see data access as a right, with only small percentages disagreeing (3% US, 1% Ireland). Given that privacy is discussed in human rights terms more in European policy discussions than in the US, we might expect that difference to be mirrored in our participants. However, both our US and Irish participants strongly agreed that they want a data access right (63% US, 68% Ireland). The difference in desire is not statistically significant ($p=.106$). In a follow up free-form text response, the most popular sentement participants reported for answering as they did was: "It is my data".

The idea of an access right is abstract and affects society as a whole. We asked a more concrete and personal question of whether participants would be interested in exercising their right to access their own data.

As shown in Fig. 4, both US and Irish participants have a strong desire to use a right of access to see data

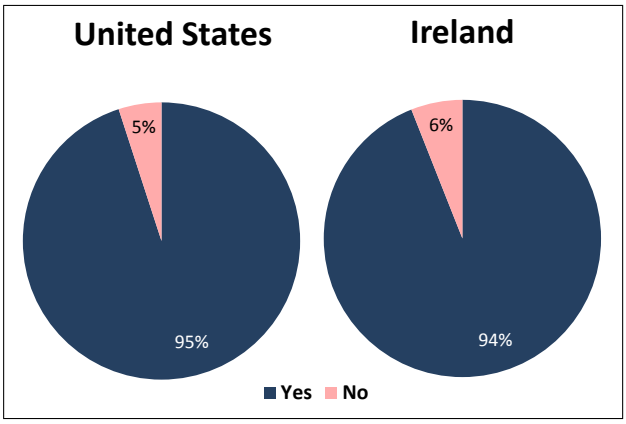

Fig. 4. Desirability to use the right to access data.

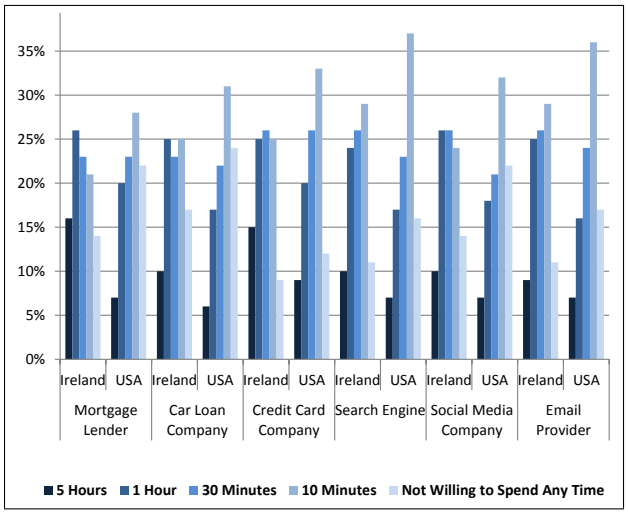

Fig. 5. Time participants are willing to spend to access data.

held about them by online companies, with no statistically significant differences. Only small percentages say they would not use access rights (5% US, 6% Irish). The main reasoning behind the desire to use the right to access for both populations was to see what data has been collected (58% US, 57% Ireland). The desire to correct inaccuracies in collected data ranked as the second most popular reason (41% US, 40% Ireland). Of the small percentage of participants that stated they weren't interested in accessing their data (5% US, 6% Ireland), disinterested US participants stated most frequently that they didn't care what data the company held about them (46%). Disinterested Irish participants were most likely to feel that the process of accessing their data would take too much of their time (48%).

We found statistically significant differences of self-reported experiences accessing data. Despite 53% of Americans correctly knowing that they can access their data through credit reporting agencies and being informed at the start of the survey that "data" includes financial information, the majority of participants report that they have never tried to access their data before (83% US, 92% Ireland; $p<.000$). It is easier for Irish Internet users to access data held about them by companies due to the protection afforded by the data protection act, yet fewer Irish participants have attempted to access their data. The main reason for both populations for never attempting to access their data was that they did not know the option existed (66% US, 38% Ireland; $p<.000$). Of the few who did access their data, over half of those participants stated that it was not easy to access their data (56% US, 51% Ireland; not significantly different with $p=0.534$). The most popular sources for accessing data among those who made an attempt were credit reporting agencies for both US and Irish participants.

Trading Time and Money for Access: We asked users to rank how much time and money they would be willing to spend to access their data from various sources. Overall, users were more willing to spend time than money on access.

As shown in Fig. 5, US participants are most willing to spend time to learn what is held about them by companies with whom they might have a financial relationship, ranking credit card companies and mortgage lenders as first and second, respectively. The same is true of our Irish participants. However, Irish participants are more likely to spend more time accessing data from mortgage lenders than credit card companies. Participants from both populations are unwilling to spend much time accessing data from online service providers. Differences in willingness to spend time to access data were statistically significant across every category of this question (overall ANOVA, $p<.000$; mortgage, $p<.000$; car loan, $p<.000$; credit card, $p=.002$; search engine, $p<.000$; social media company, $p<.000$; email, $p<.000$).

We asked a parallel question about participants' willingness to pay for access as shown in Fig. 6. The most popular amount was \$0. Differences in willingness to pay for accessing data were not statistically significant ($p=.624$). Currently, any Irish citizen who makes an access request under Section 4 of the Data Protection Act is charged a fee of €6.35 [6]. US citizens are entitled to one free credit report per company per year under the Fair and Accurate Credit Transactions Act [3] but subsequent reports cost \$7.95 as regulated by the FTC Fair Credit Reporting Act. As Fig. 6 indicates,

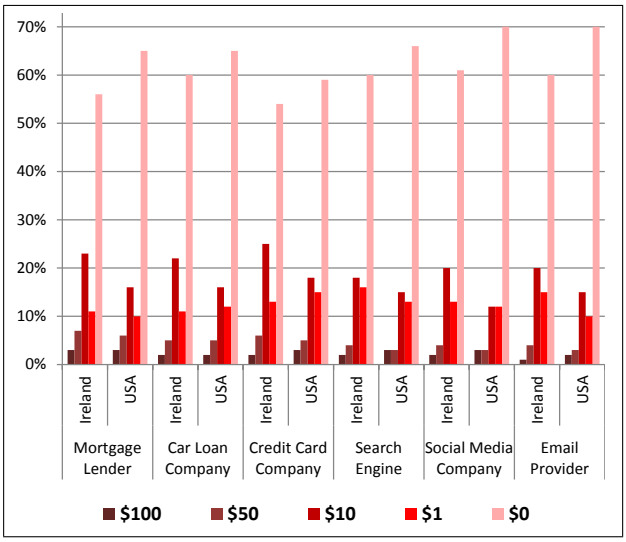

Fig. 6. Money participants are willing to spend to access data.

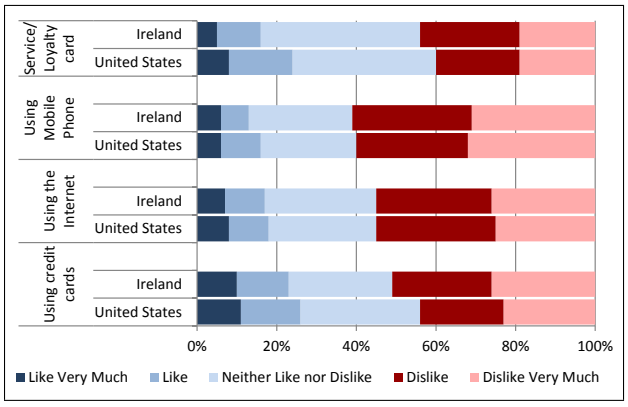

Fig. 7. Attitudes towards data being recorded to deliver services.

both our US and Irish participants were most willing to pay money to access data from companies with whom they might have a financial relationship.

6.3 Attitudes Towards Data Brokers

In this section we study the effect data brokers have on participants' attitudes to data access.

Recorded Behavior: Today, nearly every transaction Internet users engage in records their behavior for a range of reasons, such as fraud prevention or delivering ads suited to their interests. We asked users how they feel about this practice when signing up for a service/loyalty card, when using their mobile phone, when using the Internet and when using their credit cards.

As seen in Fig. 7, US participants most like data collection around credit card use, likely to prevent fraud as the question stem suggests. They are ambivalent about data collection from offline first parties, and most dislike data collection while using the Internet and mobile phones. Irish participants share similar views but a greater percentage of Americans are more receptive to their behavior being recorded to deliver services. According to the 2015 European Commission Eurobarometer on Data Protection [69], Europeans reported that they were most concerned about the recording of their activities via mobile phones. The same is true of our Irish participants.

Existing Data Brokers: Acxiom, CoreLogic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, RapLeaf and Recorded Future are some of the largest data brokers. The FTC notes that most consumers have never heard of the data broker industry, let alone the names of the largest data brokers [43]. We asked participants to name the leading data brokers in their country using a free-form text response. The majority of US and Irish participants answered with "Unsure" (83% US, 90% Ireland). The next most popular responses by US participants were Experian which is a credit reporting agency (3%) and Google (3%). The next most popular responses by Irish participants were Google (1%) and the Data Protection Commission (1%).

We also asked participants about their interest in accessing data held about them by data brokers. More than two fifths of US and Irish participants expressed strong interest (43% US, 44% Ireland), a similar number expressed slight interest (45% US, 44% Ireland) and the remaining participants reported that they were not at all interested (12% US, 12% Ireland). Despite participants being unable to name any major data brokers, that they want access to data held about them by data brokers indicates that Internet users have a desire for access more generally. There were no significant differences in interest between US and Irish participants to access data held about them by data brokers (p=.931). This suggests that both populations want access. We asked participants to indicate from a list of options the most effective method to learn about their data access rights. The most popular choice was a dedicated Internet website showing the information collected about the user (55% US, 53% Ireland). A warning by the company collecting the data in advance of entering a transaction with that company was the second most popular choice for US participants (20%) followed by periodic informative emails regarding access rights (15%) while television ads informing you of your rights ranked second for Irish participants (19%) followed by company warnings

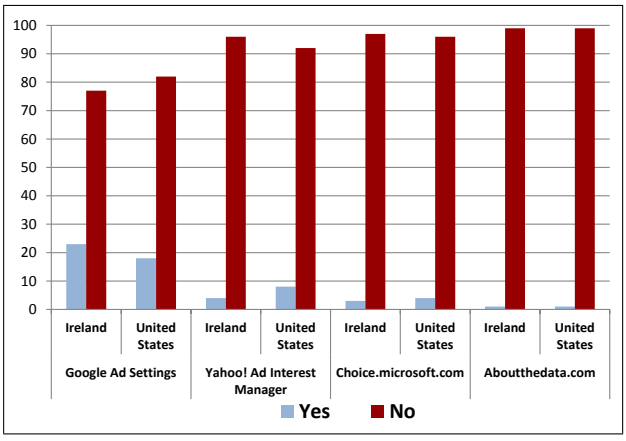

Fig. 8. Use of existing access tools.

(18%). TV ads fared as the worst method for US participants (10%) while periodic informative emails fared the worst for Irish participants (10%).

We asked participants if they have used existing access tools, such as online ad managers and data broker Acxiom's aboutthedata.com. As shown in Fig. 8, the majority of US and Irish participants have never used existing ad managers. Of those who have, the most common tool is Google Ad Settings. Only 1% of US and Irish participants reported using aboutthedata.com. When asked what these users think aboutthedata.com is used for, most of the participants reported that they didn't know or gave a general statement such as to manage or understand data. This suggests that Acxiom's attempt at transparency is not reaching Internet users.

6.4 Proposed Models

We created three separate model websites which citizens could use to access their data. Participants were shown either model A and C or B and C, but not all three. We sought to gain true insight as to whether participants are more interested in an expanded ad manager rather than a scaled down one. It may have been tempting for a participant to rate a tool which shows more information as more useful directly after seeing a scaled down tool. This strategy avoids this.

Participants were presented with mockup websites rather than actual systems as this allows us to reduce the number of variables tested at a time and ensures people are reacting to the differences between models, not to confounds. For example, if participants were asked to compare Google and Yahoo! ad managers, the results may be skewed due to a range of factors such as brand loyalty, interface design or familiarity. Creating our own systems avoids this. Additionally, the more detailed a user interface (UI) is, the more people are likely to react to the details of the UI rather than the information it conveys. By engaging in prototyping, we are able ensure participants react to the information conveyed by the systems rather than superficial elements. Amount of data was chosen as the main design variable for the models because existing access tools give limited insight to users by only showing broad categories of data rather than specific data points. We sought to examine whether a more detailed access tool than is currently available is something participants are interested in and would find useful.

We refer to the models as A, B and C for the purposes of this paper, in order of least informative to most complex. Models A (see Fig. 9) and B (see Fig. 10) represent a fictional website called AdConnections.com (inactive domain as of March 2015). This website would act as a hub where an Internet user could view which behavioral categories a company claims the user is interested in. Model A would allow a user to see what categories of ads have been assigned to them but nothing more. In Model B we add information about why a user is in a specific category. For any category, clicking a button displays an AdConnections profile revealing how the link between the behavioral category and the user occurred. For instance, an ad for Virgin America could be established because the user searched for the terms "Holiday," "Domestic Flights," "USA," and "Miami," leading to a "Travel" behavioral category. We tested these two variants to compare whether the expanded version (Model B) was overly detailed, or whether the scaled down version (Model A) was not detailed enough. Practically, it would be easier to set up a website similar to Model A, just reporting behavioral categories. Model A is similar to existing tools available today such as Google Ad Settings [40], Yahoo Ad Interest Manager [75], and choice.microsoft.com [54]. It would require much more time and money to construct a website similar to Model B, and could require co-ordination between multiple companies. A 2014 FTC report on data brokers recommended the creation of a centralized mechanism, such as an Internet portal, where data brokers can identify themselves, describe their information collection and use practices, and provide links to access tools and opt outs [35]. Model C (see Fig. 11) responds to FTC Commissioner Brill's "Reclaim Your Name" proposal to establish technical controls allowing people to access the information data brokers have stored about them, control how it is shared and correct it when nec-

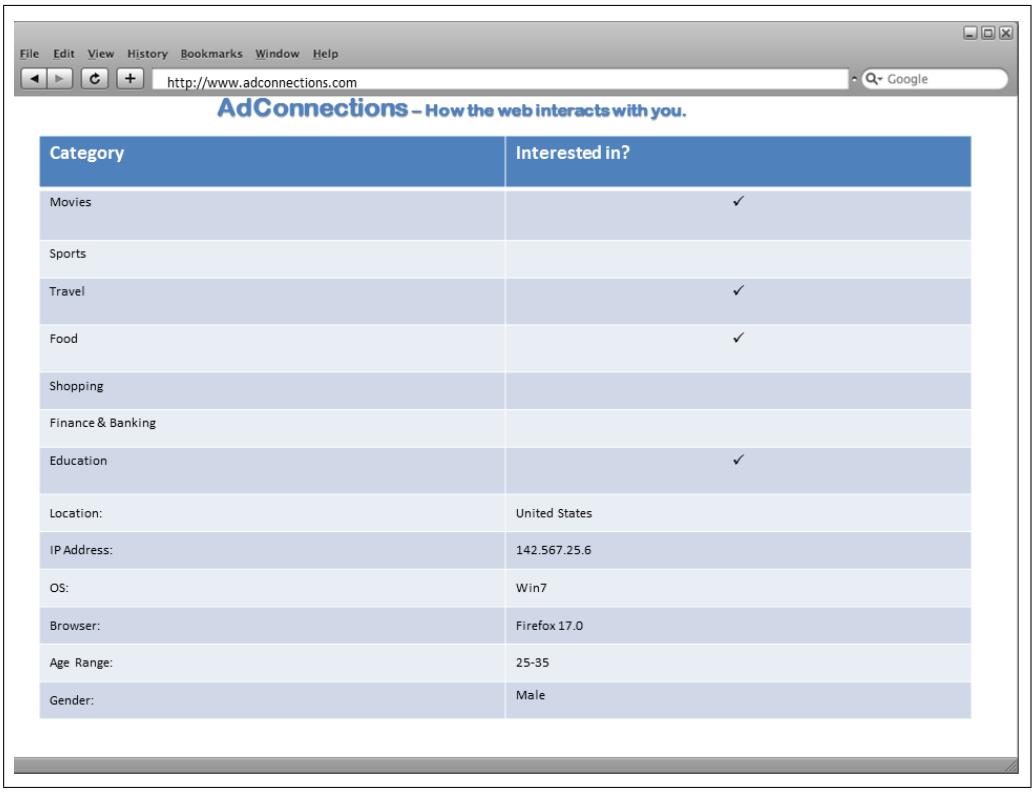

Fig. 9. Model A – Scaled down AdConnections. Shows interest categories on the left, with a checkmark for those that apply.

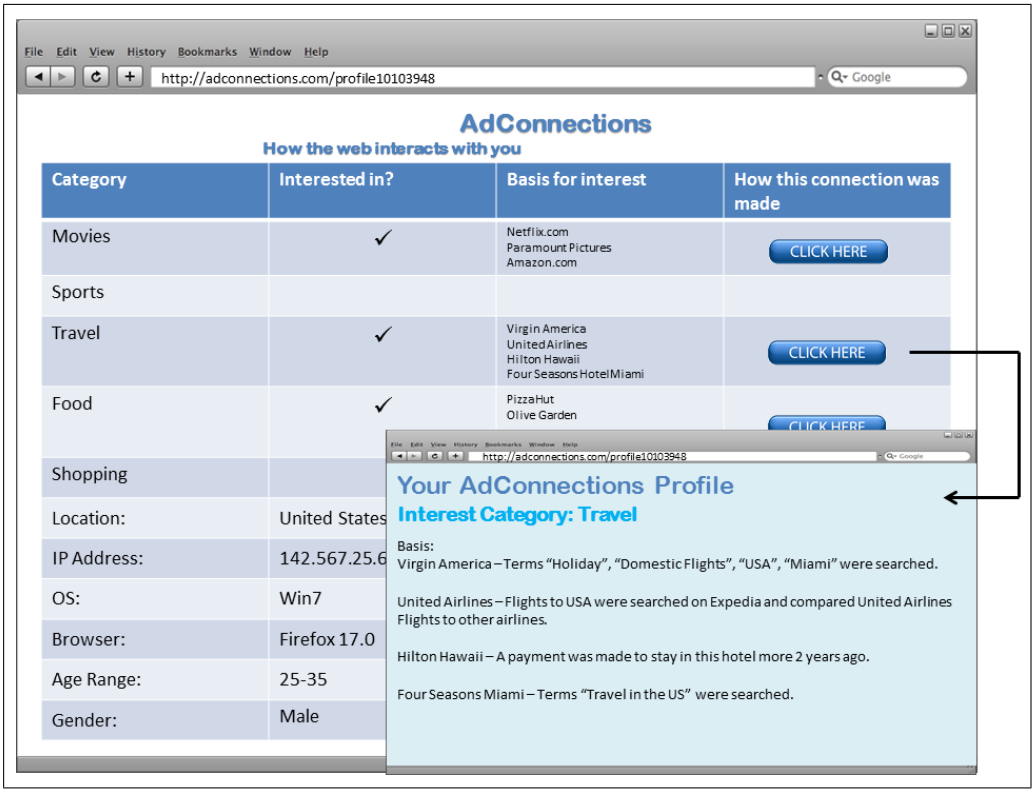

Fig. 10. Model B – Expanded AdConnections. Shows interest categories on the left and checkmarks for those that apply as in Model A, plus why interest categories apply, and an optional pop-up window for the details of category assignment.

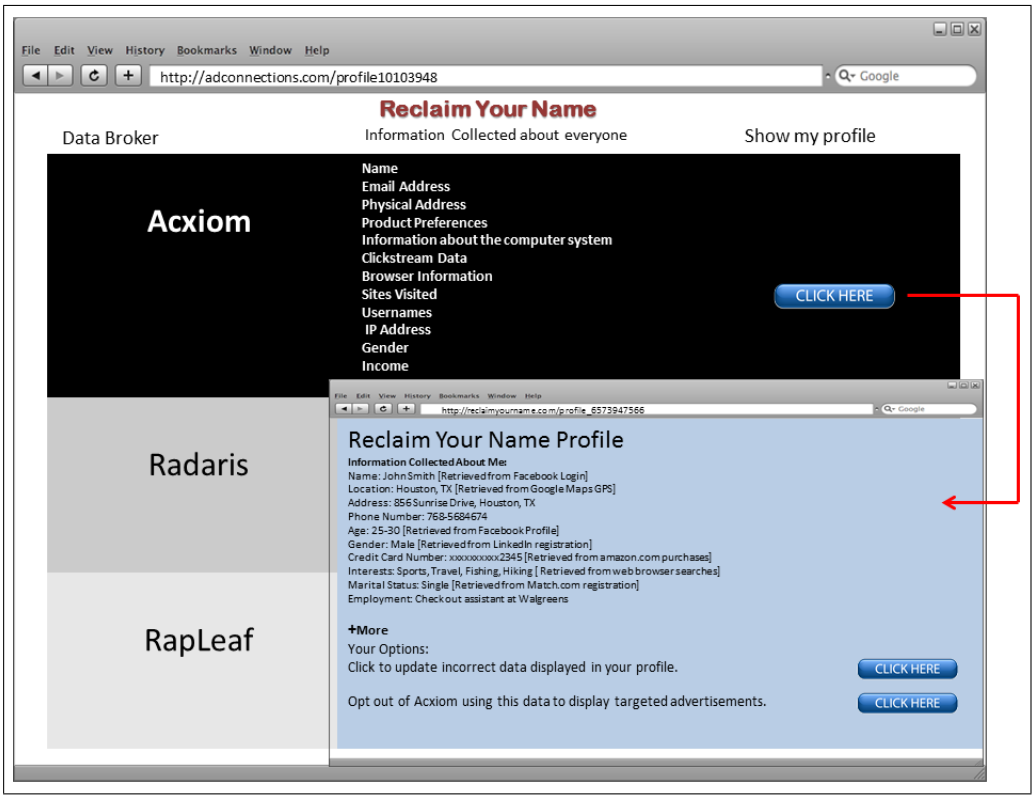

Fig. 11. Model C – Multiple Data Broker Hub (response to Reclaim Your Name). Shows the same data as in Model B, but for multiple data brokers rather than just one.

essary [13]. We envisioned a central portal with multiple data brokers listed in one place.

Participants were not informed outright that models A and B are platforms that allow them to access data from a single company and that model C allows them to access data from multiple brokers but this was ascertainable through context. Models A and B are similar to existing ad managers which are per company. An in-survey guidance note informed participants that through model C, users would be able to access information data brokers have stored about them. Model C also shows "data broker" as an interface heading and lists Acxiom, Radaris and RapLeaf as examples. This provided further indication that model C is a hub for multiple brokers.

Results from Comparing the Three Models: Across all three proposed models, the responses by both US and Irish participants were similar.

Model A, the tool with the least detail but also least complicated, fared worse than model B but performed better than model C, as shown in Fig. 12. A higher percentage of participants who were shown model B, which expands upon model A, rated it as a useful tool (71% US, 72% Ireland) compared to those shown model A (64% US, 71% Ireland). Model C, which adds the complexity of a hub for multiple data brokers, performed the worst but still did well (64% US, 60% Ireland).

The pattern above repeats when participants were asked whether they would use such tools, as shown in Fig. 13. Again, a higher percentage of participants who were shown model B were interested to use the tool (69% US, 75% Ireland) compared to those shown model A (64% US, 68% Ireland). Model C ranked last (64% US, 64% Ireland).

We asked participants to describe the level of detail they feel is included in each tool. Over half of our American and Irish participants who were shown model A feel that model A is not detailed enough (58% US, 55% Ireland). Fewer participants who were shown model B feel that model B is overly detailed (33% US, 37% Ireland). Even fewer participants feel that model C is not detailed enough (25% US, 22% Ireland). In engineering an access tool, it is important to take into account the level of detail provided by the interface. Too much information, such as that provided in model C, may overwhelm the user but too little, as in model A, defeats the purpose of access. Engineering the concepts put forward in models

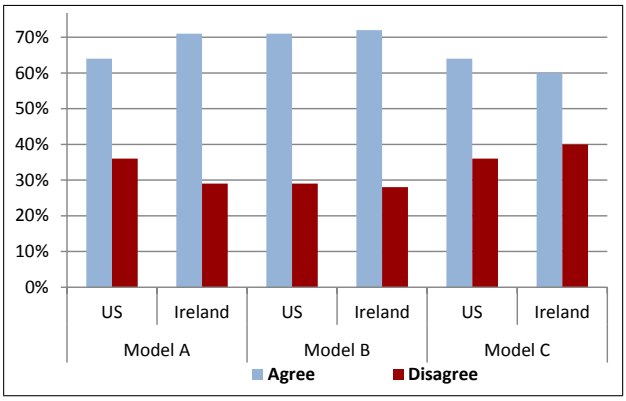

Fig. 12. Participant views on whether mockups are useful.

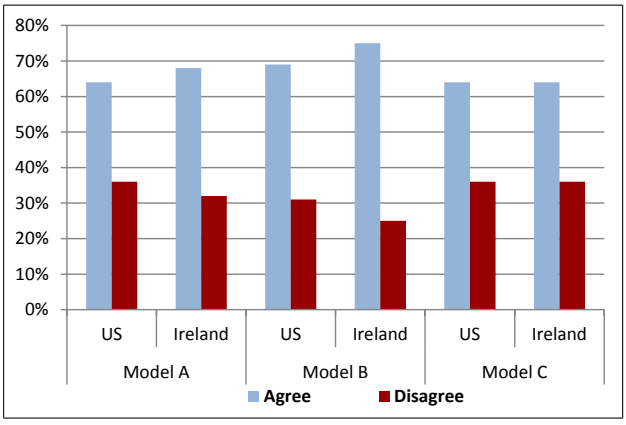

Fig. 13. Participant views on using mockups if available.

A, B and C and testing them in practice is an important area of future work regarding access.

We asked participants what they would do with information that could be gained from a multiple data broker hub. The majority of participants report that they would opt-out of targeted ads (59% US, 58% Ireland). The second most popular action participants would take is to update incorrect information that data brokers hold about them (37% US, 36% Ireland). While providing access to data and nothing more is a good first step in itself, there may be more value in access if it is paired with the ability of users to control their data through opt-out and data correction.

7 Discussion & Conclusion

While there have been numerous studies conducted on attitudes to data collection and online behavioral advertising, our study is the first to investigate attitudes about data access across two countries and to compare results. Our research:

- Establishes low awareness levels surrounding data access among Internet users;
- Demonstrates high interest levels among Internet users to access the data collected about them by online companies;
- Establishes low knowledge levels surrounding the data broker industry among Internet users;
- Prototypes policy-relevant technical tools that allow Internet users to gain greater access and control over data held about them by online companies and
- Establishes that access is important to US and Irish Internet users and existing access mechanisms are inadequate to serve the needs of either population.

Awareness: There seems to be a state of confusion among participants from both populations as to what sources they can use to access their data. Almost a fifth of American participants incorrectly report that they can access data held about them by writing to a Data Protection Commission. Nearly half of Irish participants were unsure as to whether they could access their data in this way. Only 45% of Irish participants correctly knew that they can. More Americans know of their right to access their data through credit reporting agencies (53%) but under half of their Irish counterparts know they have the same right (43%). We find that laws and bills reported in the media may lead to confusion between which laws exist in which countries. Awareness of laws may be strengthened through the practice of utilizing the laws rather than simply hearing about them.

American participants believe that the US has stronger access laws than other countries. Irish participants believe Irish laws provide the same levels of access as laws in other countries. While neither country has an overwhelming advantage, Irish Internet users are able to access their data more easily as a result of the protections of the Data Protection Act. It is likely that since participants lacked knowledge of their own laws, they were unable to assess how their laws contrast to laws in other nations. A lack of knowledge of laws may also lead to participants answering blindly in overconfidence for their legal system.

Several companies are beginning to educate their users on how they collect and process user data. Google provides a walkthrough and information prompt with the launch of their new privacy hub [50]. Users are prompted to review the privacy hub on logging in or when visiting Google's homepage. This encourages Internet users to explore the access options available to

them. In 2014, Facebook attempted to make users more aware of their privacy settings by introducing a dinosaur mascot which prompts users to review their settings [42]. In the same way, Google's walkthrough could be a good step towards raising awareness around access.

Desirability: American and Irish participants report that they want the right to access their data. A prior study by Cranor indicates that users are concerned about the security of their information [23]. If users have the right to access data, they may feel more confident knowing exactly what data is at risk of security breaches. Understanding the relationship between user trust and access rights is a promising area of future work. More American than Irish participants have tried to access their data despite a more extensive access mechanism being available in Ireland through the Data Protection Act. We found that in general, Irish participants were willing to spend more time and money to access their data . According to the 2011 Eurobarometer report, 38% of Irish citizens are prepared to pay for access [68]. Both groups of participants were more likely to self-report willingness to spend time than money to access data.

While we reaffirmed that participants are highly resistant to paying to access data held about them, they rarely know they can do so if there are few economic incentives to inform them of data access opportunities. This is borne out in US participants' consistently high familiarity with credit reporting agencies. The hybrid approach of one free credit report per company per year and additional credit reports and scores for profit may be largely disliked, but we speculate it could be the best thing to happen for consumer data access in practice. Credit reporting agencies have incentive to educate users about their access rights, as credit reporting agencies profit from access. In 2015, Experian made $980 million from providing access to credit reports [29], and Equifax made $51 million in 2014 [26].

Data Brokers: Less than 1% of US participants and none of our Irish participants were able to identify any of the major data brokers. Participants from both countries generally dislike the fact that their behavior is recorded to deliver services to them such as targeted ads. Based on these results, we find that although Internet users are largely unaware of who data brokers are, they want more transparency over the activities data brokers engage in. Transparency can be achieved by providing access. We find that while websites such as aboutthedata.com are a step in the right direction to providing transparency, there is little awareness raised as to the availability of such access options to Internet users. Only 1% of US and Irish participants report that they have used the website. In a follow up question asking for details regarding what the website is used for, participants appear to have no idea. Abouthedata.com received only 9,000 desktop visits worldwide in June 2015 [62].

Solutions and the Future: We tested two different interfaces to accessing information from online companies and one interface to accessing information held by data brokers. Participants from both populations stated that these tools would be useful to them. Participants preferred the expanded version of our mock website AdConnections.com over the scaled down, simplified version. Unfortunately, most existing tools are similar to the scaled down version. Data access tools could give users more control over who sees and uses their data, and could enhance online trust. Future work could determine if users would be more willing to share data if they could better control what happens with that data. Model C, the data broker hub, fared the worst of the three systems but still received a positive response. We find that participants may have found the other tools more understandable as users engage with online companies on a daily basis but data brokers operate in the background and users have little contact with them.

Future research could create such tools and run them in practice to gain further usability insights. We suggest that companies building new tools work to provide enhanced transparency of not just which behavioral categories they assign, but on what basis. We understand this can be non-trivial to engineer. We suggest that companies continue working on their own interfaces and educating their users on how to use them, with an expectation of a federated hub to follow.

While our research provides insights into attitudes of Internet users towards data access, more research questions arise. For example, how effective would implementing a tool such as AdConnections.com be in reality? Anecdotal reported rates of use for existing tools are exceedingly low. How do Internet users in other countries with different legal systems feel about data access? These questions await future research.

Access as a privacy protection is integral in ensuring the effectiveness of the other FIPPs. By ignoring access, the other protections are weakened. For example, without access regulators might find it more difficult to verify the authenticity of notice, and thus enforcement is weakened. Our work highlights that access is important to Internet users but they have no idea which third parties hold their data. As our research shows, Internet users haven't heard of data brokers. As a result, ask-

ing Internet users to go to each separate company that holds data about them is too high a burden for effective access. While a hub website a might be too complex to engineer at present, while it is new to users, in the long run it seems to be the only model of the three that can provide effective access to Internet users. However, any attempts at access would be welcome so we should not wait for the perfect coordinated solution.

8 Acknowledgments

This work was conducted while both authors were at Stanford University. The authors have since left Stanford but thank Barbara van Schewick and the Stanford Center for Internet and Society for making this work possible. The authors would also like to thank the PETS reviewers for their helpful comments and feedback. Finally, the authors wish to thank Claudia Diaz, one of the PETS PC Chairs this year, who was particularly generous with her time and help in the submissions process.

References

[1] 104th Congress, Public Law 104-191. Health Insurance Portability and Accountability Act of 1996, 1996.

[2] 105th Congress, Public Law 105-277. Children's Online Privacy Protection Act of 1998, 1998.

[3] 108th Congress, Public Law 108-159. Fair and Accurate Credit Transactions Act of 2003, 2003.

[4] 114th Congress, S.668. Data Broker Accountability and Transparency Act of 2015, 2015.

[5] A. Acquisti and R. Gross. *Privacy Enhancing Technologies*, chapter Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook, page 11. Springer Berlin Heidelberg, 2006. G. Danezis and P. Golle, eds.

[6] Act no. 6 of 2003. Data Protection (Amendment) Act 2003, 2003.

[7] Acxiom. Viewing and editing data about you. http://documentation.acxiom-online.com/aspect/#b6811t5921n/s-1/s15/s15b1851/s15b1859 [Accessed: 07- Jun- 2015].

[8] American Civil Liberties Union. Right to Know Act (ab 1291). https://www.aclunc.org/our-work/legislation/right-know-act-ab-1291 [Accessed: 02- Jun- 2015].

[9] J. Angwin. *Dragnet Nation: A quest for privacy, security, and freedom in a world of relentless surveillance*. Times Books, 2014.

[10] K. Bachman. Senate commerce report says data brokers 'operate behind a veil of secrecy', 2013. http://www.adweek.com/news/technology/senate-commerce-report-says-data-brokers-operate-behind-veil-secrecy-154579 [Accessed: 17- Jun- 2015].

[11] Better Regulation. EU Data Protection Regulation, 2014. http://www.betterregulation.com/ie/data-protection-proposals [Accessed: 17- Jun- 2015].

[12] M. Brantley. Acxiom begins consumer access to some data, 2013. http://www.arktimes.com/ArkansasBlog/archives/2013/09/04/acxiom-begins-consumer-access-to-some-data [Accessed: 17- Jun- 2015].

[13] J. Brill. "Reclaim Your Name" keynote address at the 23rd computers freedom and privacy conference, 2013. https://www.ftc.gov/sites/default/files/documents/public_statements/reclaim-your-name/130626computersfreedom.pdf [Accessed: 17- Jun- 2015].

[14] Business World. Irish data protection is not "soft", 2015. https://www.businessworld.ie/european-news/Irish-data-protection-is-not-soft--1667.html [Accessed: 20- Jul- 2015].

[15] California Business and Professions Code section 22575-22579, online privacy protection act, 2003. http://www.leginfo.ca.gov/cgi-bin/displaycode?section=bpc&group=22001-23000&file=22575-22579 [Accessed: 14-Feb-2016].

[16] California AB-1291. Right to Know Act of 2013: disclosure of a customer's personal information, 2013.

[17] California Civil Code section 1798.83. California Shine the Light law, 2005.

[18] Cint. Cint - about us, 2015. http://www.cint.com/about [Accessed: 06- Apr- 2015].

[19] M. Cogley. Of course Facebook would go to a country with the lowest levels of data protection, 2015. http://www.newstalk.com/Of-course-Facebook-would-go-to-a-country-with-the-lowest-levels-of-data-protection- [Accessed: 20- Jul- 2015].

[20] Administration discussion draft: Consumer Privacy Bill of Rights Act of 2015, 2015. https://www.whitehouse.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf [Accessed: 15- Jun- 2015].

[21] Council of the European Union. Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation) 9565/15, 2015.

[22] Council of the European Union. Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation) [first reading]-analysis of the final compromise text with a view to agreement 15039/15, 2015.

[23] L. Cranor, J. Reagle, and M. Ackerman. *The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy*, chapter Beyond concern: Understanding net users' attitudes about online privacy, pages 47–70. MIT Press, 2000. I. Vogelsang and B. Compaine, eds.

[24] S. Curtis. Facebook ordered to stop tracking non-users in Belgium or face fines, 2015. http://www.telegraph.co.uk/technology/facebook/11985694/Facebook-ordered-to-stop-tracking-non-users-in-Belgium-or-face-fines.html [Accessed 22-Feb-2016].

[25] P. Dixon. Testimony of Pam Dixon executive director, World Privacy Forum before the Senate Committee on Commerce, Science, and Transportation, "what information do data brokers have on consumers, and how do they use it?", 2013. http://www.worldprivacyforum.org/wp-content/uploads/

2013/12/WPF_PamDixon_CongressionalTestimony_DataBrokers_2013_fs.pdf [Accessed: 15- Jul- 2015].

[26] Equifax. Equifax 2014 annual report, 2014. http://www.equifax.com/pdfs/corp/Equifax_2014_Annual_Report.pdf [Accessed: 03- Aug- 2015].

[27] Europe Versus Facebook. Get your data! make an access request at Facebook! http://www.europe-v-facebook.org/EN/Get_your_Data_/get_your_data_.html [Accessed: 25- Oct- 2015].

[28] European Commission. Why do we need an EU data protection reform?, 2012. http://ec.europa.eu/justice/data-protection/document/review2012/factsheets/1_en.pdf [Accessed 28-Nov-2015].

[29] Experian. Experian annual report 2015, 2015. http://annualreport.experianplc.com/2015/_resources/pdf/ExperianAnnualReport2015.pdf [Accessed: 03- Aug- 2015].

[30] Federal Trade Commission. Privacy online: A report to Congress. Technical report, FTC, 1998.

[31] Federal Trade Commission. A summary of your rights under the Fair Credit Reporting Act. Technical report, FTC, 2009. https://www.consumer.ftc.gov/articles/pdf-0096-fair-credit-reporting-act.pdf [Accessed: 20-Jul-2015].

[32] Federal Trade Commission. Cookies: Leaving a trail on the web. Consumer Report, 2011. https://www.consumer.ftc.gov/articles/0042-cookies-leaving-trail-web [Accessed: 14-Feb-2016].

[33] Federal Trade Commission. FTC to study data broker industry's collection and use of consumer data, 2012.

[34] Federal Trade Commission. Protecting consumer privacy in an era of rapid change. Technical report, FTC, 2012.

[35] Federal Trade Commission. Data brokers: A call for transparency and accountability, May 2014. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf [Accessed: 29-Nov-2015].

[36] S. Fennell. Information re data brokers in Ireland. Personal email, Sept. 9 2014. e-mail: infodataprotection.ie Message: 'In relation to your email you may wish to direct any queries you have to the Irish Brokers Association. The following link will take you to their website: http://www.iba.ie/'.

[37] J. Fromholz. The European Union Data Privacy Directive. *Berkeley Technology Law Journal*, 15(1):471–472, 2000.

[38] FundingUniverse. Acxiom corporation history, 2011. http://www.fundinguniverse.com/company-histories/acxiom-corporation-history/ [Accessed: 06- Jun- 2015].

[39] S. Gibbs. Belgium takes Facebook to court over privacy breaches and user tracking, 2015. http://www.theguardian.com/technology/2015/jun/15/belgium-facebook-court-privacy-breaches-ads [Accessed: 20- Jul- 2015].

[40] Google. Control your Google ads. https://www.google.com/settings/ads/ [Accessed: 15- Jul- 2015].

[41] Google. Privacy and terms. https://www.google.com/policies/privacy/key-terms/#toc-terms-personal-info [Accessed: 14-Feb-2016].

[42] D. Gross. Forget godzilla: Facebook rolls out its own dinosaur, 2014. http://edition.cnn.com/2014/05/23/tech/social-media/facebook-dinosaur-mascot/ [Accessed: 15- Jul- 2015].

[43] G. Gross. FTC: Congress should rein in data brokers, 2014. http://www.pcworld.com/article/2168060/ftc-congress-should-rein-in-data-brokers.html [Accessed: 15- Jun- 2015].

[44] H. Teufel III. The Fair Information Practice Principles: Framework for privacy policy at the Department of Homeland Security. Technical report, The Privacy Office, U.S. Department of Homeland Security, 2008.

[45] Irish Brokers Association. About the IBA. http://iba.ie/about-us/about-the-iba/ [Accessed: 09- Sep- 2014].

[46] Irish Brokers Association. Irish brokers association. http://iba.ie [Accessed: 09- Sep- 2014].

[47] J. Jerome and B. Dambrine. Comparing the Data Broker Bill to the Consumer Privacy Bill of Rights, 2015. http://www.futureofprivacy.org/2015/03/16/comparing-the-data-broker-bill-to-the-consumer-privacy-bill-of-rights/ [Accessed: 15- Jun- 2015].

[48] K. Harris, Attorney General of California. Privacy on the go, recommendations for the mobile ecosystem, 2013. http://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/privacy_on_the_go.pdf [Accessed: 14-Feb-2016].

[49] A. Kittur, E. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. *26th Special Interest Group on Computer-Human Interaction (SIGCHI) Conference*, 2008.

[50] M. Liedtke. Google demystifies privacy controls with new redesign, 2015. http://www.inc.com/associated-press/google-tries-to-demystify-privacy-controls-with-new-approach.html [Accessed: 15- Jul- 2015].

[51] K. Lillington. Strong data protection laws better for EU than sniping, 2015. http://www.irishtimes.com/business/technology/strong-data-protection-laws-better-for-eu-than-sniping-1.2185370 [Accessed: 20- Jul- 2015].

[52] M. Madden and L. Rainie. Americans' attitudes about privacy, security and surveillance, 2015. http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/ [Accessed: 29-Nov-2015].

[53] Microsoft. Safety & security center. https://www.microsoft.com/security/online-privacy/overview.aspx [Accessed: 15- Jul- 2015].

[54] Microsoft. Your privacy and Microsoft personalized ads. http://choice.microsoft.com/en-US [Accessed: 15- Jul- 2015].

[55] P. Newenham. Facebook responds to Belgian tracking claims, 2015. http://www.irishtimes.com/business/technology/facebook-responds-to-belgian-tracking-claims-1.2219799 [Accessed: 20- Jul- 2015].

[56] L. Newman. Here's how Facebook chooses which ads to show you, 2014. http://www.slate.com/blogs/future_tense/2014/08/14/facebook_s_why_am_i_seeing_this_shows_what_the_company_knows_about_you.html [Accessed: 10- Jun- 2015].

[57] J. O'Connor and A. Bohan. *Getting the Deal Through – Data Protection & Privacy*, chapter Ireland, pages 73–81. Gideon Roberton, 2014. Rosemary P Jay, ed. http://www.matheson.com/images/uploads/documents/Ireland_GTDT_Data_Protection__Privacy_2014.pdf [Accessed: 17- Jun- 2015].

[58] Official Journal of the European Union, L 281. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free

movement of such data, 1995.
[59] Organization for Economic Co-operation and Development. Guidelines on the protection of privacy and transborder flows of personal data. Technical report, OECD, 1980.
[60] Pew Research Center. Internet user demographics, 2014. http://www.pewinternet.org/data-trend/internet-use/latest-stats/ [Accessed 30-Nov-2015].
[61] M. Scott. Facebook to appeal a Belgian court's ruling on data privacy, 2015. http://www.nytimes.com/2015/11/11/business/international/facebook-belgium-privacy.html [Accessed 22-Feb-2016].
[62] SimilarWeb. Aboutthedata.com traffic overview, 2015. http://www.similarweb.com/website/aboutthedata.com [Accessed: 15- Jul- 2015].
[63] N. Singer. Mapping, and sharing, the consumer genome, 2012. http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html [Accessed: 17- Jun- 2015].
[64] O. Solon. How much data did Facebook have on one man? 1,200 pages of data in 57 categories, 2012. http://www.wired.co.uk/magazine/archive/2012/12/start/privacy-versus-facebook [Accessed: 25- Oct- 2015].
[65] D. Solove. Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review*, 126(1880), 2013.
[66] L. Sotto and A. Simpson. *Getting the Deal Through – Data Protection & Privacy*, chapter United States, pages 191–204. Gideon Roberton, 2014. Rosemary P Jay, ed. https://www.hunton.com/files/Publication/1f767bed-fe08-42bf-94e0-0bd03bf8b74b/Presentation/PublicationAttachment/b167028d-1065-4899-87a9-125700da0133/United_States_GTDT_Data_Protection_and_Privacy_2014.pdf [Accessed: 17- Jun- 2015].
[67] M. Taddicken. The 'privacy paradox' in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. *Journal of Computer Mediated Communication*, 19(2):248, 2013.
[68] TNS Opinion & Social. Attitudes on data protection and electronic identity in the European Union, 2011. http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf [Accessed: 29-Nov-2015].
[69] TNS Opinion & Social. Data protection, 2015. http://ec.europa.eu/public_opinion/archives/ebs/ebs_431_sum_en.pdf [Accessed: 29-Nov-2015].
[70] A. Toth. Testimony of Anne Toth, vice president of policy and head of privacy , Yahoo! inc. before the Joint Hearing of the Subcommittee on Communications, Technology and the Internet and the Subcommittee on Commerce, Trade and Consumer Protection of the Energy and Commerce Committee of the United States House of Representatives on behavioral advertising: Industry practice and consumers' expectations, 2009. http://democrats.energycommerce.house.gov/sites/default/files/documents/Testimony-Toth-CTCP-CTI-Behavioral-Advertising-Practices-2009-6-18.pdf [Accessed: 15- Jul- 2015].
[71] B. Ur, P. Leon, L. Cranor, R. Shay, and Y. Wang. Smart, useful, scary, creepy: perceptions of online behavioral advertising. *Proceedings of the Eighth Symposium on Usable Privacy and Security - SOUPS*, 2012.
[72] US Senate Committee on Commerce, Science, and Transportation. A review of the data broker industry: Collection, use, and sale of consumer data for marketing purposes. Technical report, United States Senate, 2013.
[73] T. Walker. Max Schrems: The Austrian law graduate who became a champion of Facebook users, 2015. http://www.independent.co.uk/life-style/gadgets-and-tech/news/max-schrems-the-austrian-law-graduate-who-became-a-champion-of-facebook-users-a6683711.html [Accessed: 11- Oct- 2015].
[74] W. Ware. Records, computers and the rights of citizens. Technical report, Rand Corporation, Santa Monica, 1973.
[75] Yahoo! Ad interest manager. http://info.yahoo.com/privacy/us/yahoo/opt_out/targeting/details.html [Accessed: 15- Jul- 2015].
[76] Yahoo! Yahoo privacy center (ie). https://policies.yahoo.com/ie/en/yahoo/privacy/index.htm [Accessed: 11- Jun- 2015].
[77] Yahoo! Yahoo privacy center (us). https://policies.yahoo.com/us/en/yahoo/privacy/index.htm [Accessed: 11- Jun- 2015].

9 Appendix A: Survey Questions

We defined "data access," "data brokers" and what types of "data" we are asking about at the start of the survey (see Section 5.3).

9.1 Survey Questions

1. **Please indicate your gender: A.** Male, **B.** Female
2. **Please indicate your age range: A.** 18-24, **B.** 25-34, **C.** 35-54, **D.** 55+
3. **Can you access data held about you through the following?:** (1) Writing to your country's credit reporting agency, (2) Writing to your country's data protection Commission, (3) By installing the program Microsoft Access, (4) By writing to the general contact information on a company's website. **Answer Options A.** Yes, **B.** No, **C.** Unsure.
4. **Which of the following sentences closely represents your view:** Data access rights in my country allow for access to **A.** more, **B.** less, **C.** the same, types of information held by companies about me than in other countries.
5. **Name a law which exists to protect your right to access data held about you by companies:** [Text Response]

6. **I want the right to access data that a company holds about me: A.** Strongly Agree, **B.** Agree, **C.** Disagree, **D.** Strongly Disagree.
7. **Why do you feel this way?** [Text Response]
8. **If you had the right to access the data that a company holds about you, would you be interested in exercising that right? A.** Yes **B.** No.
9. **[If 8 = Yes] Why would you be interested in exercising your right to access data held about you by companies? A.** To see what data they have collected about me, **B.** To correct any inaccuracies that they have made about me, **C.** To give the company more information about me, **D.** Other.
10. **[If 8 = No] Why would you not be interested in exercising your right to access data held about you by companies? A.** I do not care what data the company has about me, **B.** The process would take too much time, **C.** Other.
11. **Have you ever tried to access your data? A.** Yes, **B.** No
12. **[If 11 = yes, Q.12-15 followed] Was it easy to access your data? A.** Yes, **B.** No.
13. **Which company did you go through to access your data?** [Text Response]
14. **Why did you try to access your data? A.** I wanted to see my credit score, **B.** I wanted to see what information companies knew about me, **C.** Other.
15. **What happened? A.** I was able to access my data easily, **B.** I was able to access my data after a long and difficult process, **C.** I was unable to access my data, **D.** Other.
16. **[If 11 = No] Why have you never tried to access your data? A.** I knew the option existed but didn't know how to access my data, **B.** I didn't desire to access my data, **C.** I didn't know the option existed, **D.** Other.
17. **What is the most you would be willing to pay to learn what data is held about you by the following types of companies? Check the highest amount you would be willing to pay per company.** (1) Mortgage Lender, (2) Car Loan Company, (3) Credit Card Company, (4) Search Engine, (5) Social Media Company, (6) Email Provider. **Answer Options A.** $100, **B.** $50, **C.** $10, **D.** $1, **E.** $0).
18. **What is the longest amount of time you would be willing to spend to learn what data is held about you by the following types of companies? Check the highest amount of time you would be willing to spend per company.** (1) Mortgage Lender, (2) Car Loan Company, (3) Credit Card Company, (4) Search Engine, (5) Social Media Company, (6) Email Provider. **Answer Options A.** 5 Hours, **B.** 1 Hour, **C.** 30 Minutes, **D.** 10 Minutes, **E.** No Time).
19. **Today, nearly every transaction you engage in records your behavior, for a range of reasons such as fraud prevention or delivering ads suited to your interests. How do you feel about this practice when engaging in the following activities?** (1) When using your credit cards, (2) When using the Internet, (3) When using your mobile phone, (4) When joining a service/signing up for a loyalty card. **Answer Options A.** Like very much, **B.** Like **C.** Neither like nor dislike, **D.** Dislike, **E.** Dislike very much.
20. **Please name the five leading data brokers in your country. If you do not know some or all please enter "unsure" in the boxes below.** [Text Response]
21. **How interested are you in accessing data held about you by data brokers? A.** Very interested, **B.** Slightly interested, **C.** Not at all interested.
22. **Why do you feel this way?** [Text Response]
23. **How likely would you be to opt out of marketing lists if you had the right to do so? A.** Very likely, **B.** Likely, **C.** Unlikely, **D.** Very Unlikely.
24. **[If 23 = Very likely/ Likely] Why do you feel this way? A.** Because targeted ads are annoying and distracting, **B.** Because targeted ads are an invasion of my privacy, **C.** Other.
25. **[If 23 = Unlikely/Very Unlikely] Why do you feel this way? A.** I don't care if ads are targeted at me, **B.** I like ads to be customized to my interests **C.** Other.
26. **Have you ever used the following?** (1) Google Ad Settings, (2) Yahoo! Ad Interest Manager, (3) choice.microsoft.com. **Answer Options A.** Yes, **B.** No.
27. **What are these tools used for in your opinion?** [Text Response]
28. **Have you ever used the website aboutthedata.com? A.** Yes, **B.** No.
29. **[If 28 = Yes] What is this site used for in your opinion?** [Text Response]
30. **What is the most effective way for you to learn about your data access rights? A.** A dedicated Internet website which shows you the information collected about you, **B.** By companies warn-

ing you of the effects that entering a transaction with them will have on your access rights, **C.** Periodic emails informing you of your data access rights, **D.** TV ads advising you of your data access rights.

31. **Please state whether you agree or disagree with the following with reference to the image.** [Participants shown Fig. 9] (1) A tool like this would be useful to me. (2) There is not enough detail in this tool. (3) If I had the chance to use a tool like this I would. (4) This tool does not provide any useful information. **Answer Options A.** Agree, **B.** Disagree.
32. **What would you do with the information that can be gained with this tool? A.** Check if the connections made to me are accurate, **B.** Not sure at this time, **C.** Other.
33. **Please state whether you agree or disagree with the following with reference to the image.** [Participants show Fig. 10] (1) I understand what this platform is trying to do. (2) If this tool was available to me I would use it. (3) There is too much detail in this tool. (4) The information in this tool is not useful to me. **Answer Options A.** Agree, **B.** Disagree.
34. **What would you do with the information that can be gained with this tool? A.** Check if the connections made to me are accurate, **B.** Find out which companies are targeting ads at me, **C.** Not sure at this time, **D.** Other.
35. **Please state whether you agree or disagree with the following with reference to the image.** [Participants shown Fig. 11] (1) I understand what this tool is trying to do. (2) This tool would be useful to me. (3) This tool is too complicated. (4) This tool is not detailed enough. (5) If I had the chance to use this tool I would. **Answer Options A.** Agree, **B.** Disagree.
36. **What would you do with the information that can be gained from this tool? A.** Update the incorrect information that Data Brokers hold about me, **B.** Opt-out of ads being targeted at me, **C.** Other.
37. **Any other comments?** [Text Response]
38. **How do you rate your expertise with computers? A.** Very High (Expert with at least one computer language), **B.** Moderately High (Above average knowledge of computers with the ability to use many programs and functions), **C.** Moderately Low (Average knowledge of computers with some knowledge of using different programs), **D.** Very Low (Knowledge of the basics to navigate the web but nothing more).
39. **Do you have children at home? A.** Yes, **B.** No.
40. **What is your marital status? A.** Single, **B.** Married, **C.** Cohabiting, **D.** Divorced, **E.** Widowed.
41. **Which of the following best describes your occupation? 1.** Accounting/Finance/Banking, **2.** Administration/Clerical/Reception, **3.** Advertisement/PR, **4.** Architecture/Design, **5.** Arts/Leisure/Entertainment, **6.** Beauty/Fashion, **G.**) Buying/Purchasing **7.** Construction, **8.** Consulting, **9.** Customer Service, **10.** Distribution, **11.** Education, **12.** Health Care, (Physical & Mental), **13.** Human resources management, **14.** Management (Senior/Corporate), **15.** Military, **16.** N/A - Unemployed, **17.** N/A - Retired, **18.** N/A - Homemaker/Housewife or Househusband, **19.** News/Information, **20.** Operations/Logistics, **21.** Other:[Text Response], **22.** Planning (Meeting, Events, etc.), **23.** Production, **24.** Real Estate, **25.** Research, **26.** Restaurant/Food service, **27.** Sales/Marketing, **28.** Science/Technology/Programming, **29.** Social service, **30.** Student.
42. **Please indicate your country? A.** United State of America, **B.** Ireland.
43. **[If 42 = United States of America] Please indicate your state?** Participant shown dropdown box of 50 states and asked to select one.
44. **[If 42 = Ireland] Please indicate your county?** Participant shown dropdown box of 32 counties and asked to select one.
45. **[If 42 = United States of America] What is your income level? A.** Less than $12,000, **B.** $12,000 - $25,000, **C.** $25,000 - $50,000, **D.** $50,000 - $100,000, **E.** More than $100,000.
46. **[If 42 = Ireland] What is your income level? A.** Less than €12,000, **B.** €12,000 - €25,000, **C.** €25,000 - €50,000, **D.** €50,000 - €100,000, **E.** More than €100,000.
47. **[If 42 = United States of America] What is the highest level of education you have completed? A.** Did Not Complete High School, **B.** High School/GED, **C.** Bachelor's Degree, **D.** Master's Degree, **E.** Advanced Graduate work or Ph.D.
48. **[If 42 = Ireland] What is the highest level of education you have completed? A.** Did Not Complete Secondary School, **B.** Secondary School (Leaving Certificate), **C.** Bachelor's Degree, **D.** Master's Degree, **E.** Advanced Graduate work or Ph.D.

9.2 General Demographics

Our US participants were 46% male; our Irish participants were 49% male, showing a balanced sample of both sexes in both nations.

Age ranges were similar across national groups. 26% of our US participants were between ages 18-34, contrasted to 29% of Irish participants. 39% of our US participants were between ages 35-54, contrasted to 38% of Irish participants. 35% of our US participants were 55 or older, contrasted to 33% of Irish participants.

Our sample population was skewed to have more education than the US Internet population as a whole, and likely shares the same skew for Irish participants. While 75% of US Internet users report a high school education or less [60], in our sample population only 45% of US participants stopped at high school, and only 52% of Irish participants. Only 3% of US Internet users continue education beyond college [60], yet our sample population had 14% continuing past college in the US, and 13% in Ireland.

Similarly, our US sample population skewed to a higher income level than the US Internet population, and it is likely our Irish sample population is likewise over-representing higher earners. 85% of US Internet users report income under $50,000 [60], while in our US sample population 58% report income under $50,000.

We would need further study to be sure, but we would expect better educated participants to be more informed about access than the general population. Our study may be slightly overly optimistic as a result, which only strengthens our conclusion that Internet users are poorly informed about their access rights.

10 Appendix B: Google Access

Google is one of several companies to offer access to information about data categories. Fig. 14 and Fig. 15 show their approach.

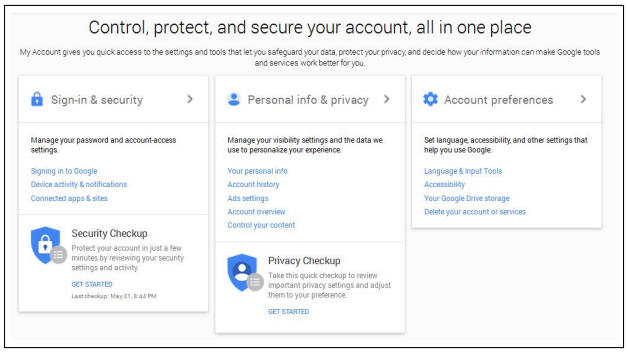

Fig. 14. Google Privacy Hub Interface.

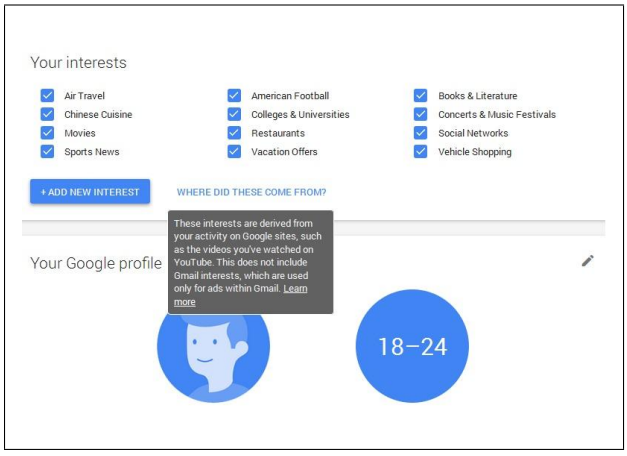

Fig. 15. Google categories.

11 Appendix C: Statistics Notes

We investigated potential statistically significant differences between our US and Irish participants (e.g., our explanatory variable is nationality). Our null hypothesis is that there is no difference between participants' nationality (H_0: $\mu_{US} = \mu_{Irish}$) tested at the 95% confidence interval ($\alpha = .05$). We had multi-part questions that were not independent so we used ANOVA to adjust the significance threshold appropriately. As Excel does not support unbalanced two-factor ANOVA tests, we used the Real Statistics plugin[5]. Unbalanced two-factor ANOVA establishes if we cannot reject the null hypothesis overall for a question, which we then followed with Chi Squared tests to determine which specific portions of a multi-part question were significantly different.

5 For an in-depth discussion of this tool, please see http://www.real-statistics.com/multiple-regression/unbalanced-factorial-anova/