

Debates, awareness, and projects about GDPR and data protection

Interim Report for the Information Commissioner's Office for the
project: *"Making new privacy rights protect and enable people's
financial futures"*

Javier Ruiz and Ed Johnson-Williams
Open Rights Group
June 2018



Table of Contents

Table of Contents	2
Introduction	4
Literature review of debates regarding the exercise of data rights	5
Right of access	5
Third parties	5
Coordination of subject access requests	6
Privacy protection hindering access	6
Right to object	7
Legitimate interests	7
Automated objection in online services	8
Right to explanation	8
To what extent does a right to explanation exist in GDPR?	8
Machine learning and the transparency fallacy	11
Right to portability	12
Right to erasure / right to be forgotten	14
Research Interviews	15
Methodology	15
Research Themes	15
Research Findings	16
Data categories and their relative sensitivity	16
Conclusions and future research	17
How organisations should behave when collecting and using data	18
Conclusions and future research	19
Control over data and GDPR emails	20
Conclusions and future research	21
Communicating about data protection and GDPR rights	22
Conclusions and future research	23
Complaining about data protection issues	23
Conclusions and future research	24
Engaging data protection rights	24
Conclusions and future research	25
Vox pops about Data Rights Finder by Project by IF	26
Findings	27
Survey of data rights tools	28
PersonalData.io	29
My Data Request	30
Philippines: Know Your Data Privacy Rights	31

Open Schufa / Selbstauskunft	31
My Data Done Right – Knowing what's known about you	33
Access My Info	34
DataRights.me	35
Terms of Service: Didn't Read	36
PriboT / Polisis	37

Introduction

This report complements the launch of the Digital Rights Finder tool delivered by Projects by IF and Open Rights Group. It is divided in three independent sections that can be read separately:

- A survey of tools for supporting data rights.
- Qualitative research on the knowledge of and attitudes to data rights among a small sample of ordinary citizens in the UK.
- A review and discussion of some of the current issues and debates around data rights under General Data Protection Regulation (GDPR).

We hope that this report will be useful for a variety of audiences. Anyone interested in technology and developing tools will find useful the only survey that collects information about the best known of these projects in a single place, including several that are still under development but will be launched soon.

Our research on data protection will be useful as a very limited snapshot of the understandings of data protection at a key moment in time, right after the GDPR had come into force.

The review on issues and potential conflicts around data rights makes accessible some of the cutting edge debates in this area, including an important discussion on the existence of a right to explanation in GDPR.

This paper will also be a very useful guidance for the next stage of the Data Rights Finder project. For example, looking systematically at other related projects gives us a good idea of what would be the most useful contributions that our efforts can make to the wider landscape.

This interim report will be followed by a final report at the end of 2018 that will outline the final research findings that guided development and the use of the tool and recommendations on best practice for the implementation of the relevant rights, which will be used to raise awareness among policy makers and industry.

We wish to thank the Information Commissioner's Office for their support in this project, and more widely for their efforts to develop the capacities of civil society to improve data protection through their grants programme.

Literature review of debates regarding the exercise of data rights

This sections provides a short literature review of relevant discussions about the data protection rights engaged in the project.

Right of access

The right of access to personal information, codified in Article 15 of GDPR, is one of the key elements of the European data protection framework, enabling the exercise of other rights and providing for the fundamental “right to data”. This right is set to expand, as GDPR removes several practical barriers to its exercise, such as the payment of fees. This new situation presents some potential challenges and controversies.

Third parties

The use of third parties for subject access requests (SARs) is fairly common, for example solicitors acting on behalf of their clients. The removal of the associated fees in GDPR will almost certainly trigger a huge increase in the use of bulk third party SARs as part of a growing contestation of data practices.

Many of the tools we discuss in this interim report facilitate SARs in ways that do not require the third party to make the request. For example, by creating templates or pre-populated emails that are sent by the data subject from their own email client. In these cases, there is no real third party, although the facilitators will bear some responsibility if the texts they provide are inaccurate or somehow lead to the request to fail.

In other cases, the intermediary will communicate directly with the organisation holding the data. This is perfectly admissible and the ICO has provided guidance on the matter¹ that essentially says that it is the responsibility of the third party to demonstrate there are entitled to act on behalf of the data subject. The ICO also says that if a data controller is concerned that the subject may not fully understand the implications of sharing the data with a third party they can send the data to the subject directly.

Organisations carrying out SARs will need to ensure they document their entitlement and that the people on whose behalf they act are fully aware of the implications, including what these organisations may want to do with the data. In some cases these third parties will want to analyse the responses for research or other purposes, or the SARs may be part of some broader complaint or legal action. This will create a new set of data protection obligations for the third party.

1

<https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/>

SARs involving children require particular care, as in principle the child should be the recipient of the response if he/she is mature enough - which can be complicated to assess.

Coordination of subject access requests

There are many projects that attempt to coordinate subject access requests targeting a company or a sector. There are concerns among some privacy activists that this could be used by some data controllers to reject the requests as excessive or manifestly unfounded, or attempt to charge a fee.

In principle each request should be considered independently, and the organisation will have to demonstrate the grounds for rejection. Batch requests are fairly common and should not be considered excessive.

The debate centres on whether a company can try to reject a coordinated SAR campaign as unfounded if they can argue that the individuals are using the SARs as a punitive tool or for reasons unrelated to data protection, for example in order to reverse engineer a database.

Recital 63 GDPR states that the right of access is there “in order to be aware of, and verify, the lawfulness of the processing”, which could be understood in fairly narrow terms of data protection. However, Art 15 GDPR simply states that individuals have the right to obtain information without any consideration as to the purposes. Given that recitals are not legally binding, it seems that there are no strong grounds for such rejection, but national courts may take a different view.

Repeated requests by the same person are a different matter, and may be considered excessive more easily if not enough time has passed or it is unlikely that enough has changed to deserve another request.

Privacy protection hindering access

One potential pitfall and controversial topic in the right of access is the extent to which privacy protecting practices may hinder the exercise of this and other data rights. Companies increasingly use pseudonymisation, data compartmentalisation and other technical measures that can make it harder to exploit the data if there were any security breaches. In some cases not even company employees can fully access the data and link it to a named individual.

These practices generally fall under the rubric of privacy – or data protection – by design, which is part of GDPR and something that normally is perceived in positive terms. The problems arise when the person trying to access the data is not a hostile third party, but the person whose data was processed in the first place.

Michael Veale, Reuben Binns and Jef Ausloos have argued that these privacy by design

techniques focus exclusively on protecting confidentiality and the identification of individuals, but the data is still potentially re-identifiable by third parties with enough capabilities. At the same time the added difficulties in identifying specific individuals make it very difficult to exercise data subject rights, such as access, erasure and objection.²

The authors document their own research with two case studies. In one case involving wifi data collected by TfL in the London Underground and used to track movements, subject access requests could not work because the data had been masked using cryptographic techniques. However, it has been demonstrated that location traces are so unique that re-identification is very easy.³

Michael Veale, Reuben Binns and Jef Ausloos also attempted to obtain recordings from the personal assistant provided in Apple products, Siri. Apple said they were unable to provide the recordings they hold because these cannot be linked to the individual, as they have different specific identifiers and they have not retrieval mechanisms. The authors made various proposals for how information systems could be engineered to improve rights while preserving privacy and who to manage any trade offs involved.

A similar case has been documented by Paul Olivier Dehaye, who asked Facebook for certain information held on him and was rejected because the date in question was kept in a backup storage and was not accessible.⁴

Right to object

Legitimate interests

The right to object to processing is important although somewhat limited in practice. The main limitations are when the data is processed under some form of official authority or public interest task, normally by the public sector, and when organisations claim a legitimate interest as the basis.

This latter point is particularly problematic, as there is little understanding among the public to the power that this gives to companies to handle their data. Legitimate interest even extends to third parties, not just the organisation processing the data, and could include direct marketing, in addition to any uses of data that are necessary – fraud prevention – or reasonably expected. Facebook, for example, uses legitimate interests to process their uses data for marketing purposes and will not offer a consent choice over this activity.

² Veale, M., Binns, R., & Ausloos, J. (2018). When data protection by design and data subject rights clash. *International Data Privacy Law*. <http://doi.org/10.1093/idpl/ipy002>

³ de Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M. & Blondel, V.D. Unique in the Crowd: The privacy bounds of human mobility. *Nature srep*. 3, 1376; DOI:10.1038/srep01376 (2013).

⁴ Additional evidence submitted to UK Parliament DCMS committee http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/fake-news/written/80117.html#_ftn3

Legitimate interests have to be balanced with the interests and freedoms of the people whose data is handled. In theory this should happen when the organisation decides to process the data, but in many cases this balancing will only take place if the practice is challenged under an objection. Despite what the GDPR says about the controller having to demonstrate its interests override those on the subject, this partially shifts the burden on the individual. In most cases they will need to provide counterarguments to explain how their interests and freedoms are not being respected. This can be trickier than it seems, as there are many data issues that cause discomfort and cause systemic effects but do not have a smoking gun of harm against the interests and freedoms of a specific data subject.

Automated objection in online services

Another issue in the practical implementation of the right to object is the implementation of automated objection for online services. Art 21 (5) GDPR says clearly that *“in the context of the use of information society services, and notwithstanding Directive 2002/58/EC, the data subject may exercise his or her right to object by automated means using technical specifications”*.

This provision would seem to point at technical systems such as Do Not Track, a browser setting that signals to websites not to track the user through cookies or other means. However, the provision is not well understood, to the point that even the ICO guidance about the right to object fails to mention it.

Right to explanation

Concerns about the pervasiveness of automated decision making systems in the everyday lives of citizens in developed economies have grown into a wave of anxiety about a *“black box society”*, including areas critical to our security and wellbeing, such as the financial system.⁵

Data protection and the rights in GDPR attempt to deal with some of these anxieties, but there is a great deal of controversy over the extent and effectiveness of data protection in opening the black box to provide an explanation to the functioning of automated systems.

To what extent does a right to explanation exist in GDPR?

There is a great deal of controversy over the extent that GDPR gives data subjects a right to an explanation over data driven decisions that affects them. This debate has critical implications for companies trying to implement best practice in data rights and for individuals trying to exercise these.

Oxford academics Bryce Goodman and Seth Flaxman created an uproar with their 2016 article that claimed that GDPR *“effectively creates a “right to explanation,”* whereby a user

⁵ Pasquale, F. (2015). *The Black Box Society*. Harvard University Press.

can ask for an explanation of an algorithmic decision that was made about them”.⁶

The claim centres on GDPR Article 22’s provisions on “*Automated individual decision-making, including profiling*”. They stated that the right could overhaul and potentially made unlawful many algorithms used in advertising and recommendation, but also credit and insurance risk assessments.

The effects of the measure would be far reaching when specially protected sensitive data (health, race, religion, etc.) is used, and there is a risk of discrimination. This is particularly the case if we take a broad interpretation that includes not only this data but any other data that can be used as a proxy, being correlated with special categories of sensitive data.

Goodman and Flaxman argue that without “*human-intelligible explanations of algorithmic decision making*” we simply cannot be sure that discrimination is not hidden under complex correlations, particularly with groups that may be underrepresented, such as ethnic minorities.

In response to the question of what does it mean, and what is required, to explain an algorithm’s decision they propose to answer questions such as: “Is the model more or less likely to recommend a loan if the applicant is a minority? Which features play the largest role in prediction? “

In a direct rebuttal to Goodman and Flaxman, Sandra Wachter, Brent Mittelstadt and Luciano Floridi have argued quite cogently and in detail against both “the legal existence and the feasibility of such a right” in GDPR.⁷

The authors see why a right to explanation is perceived as a key element for accountability and transparency in algorithms and automated systems but point that GDPR at best only provides a limited ‘right to be informed’ about automated decisions, not a full explanation.

Before analysing the provisions in GDPR the authors break down the concept of explanation into two main classifications. On the level of detail, there are explanations about general system functionality. These are different from explanations about specific decisions, that will focus on individual circumstances, such as the weighting of features.

The other main categorisation of explanations would be those that are given either before a decision has been made or afterwards. If a decision has not yet been made, clearly the prior explanation will have to rely on rules, but could also involve hypothetical testing.

The article tackles the three avenues for such right within GDPR in turn. Article 22 – and Recital 71 – provide for a right not to be subject to automated decision-making and give

⁶ Goodman, B., & Flaxman, S. (2016, June 28). European Union regulations on algorithmic decision-making and a “right to explanation.” arXiv.org.

⁷ Wachter, S., Mittelstadt, B., & Floridi, L. (2016, December 28). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, 2017. Available at SSRN: <https://ssrn.com/abstract=2903469>

some additional safeguards. Article 22 only gives data subjects a right to obtain human intervention, express views, or contest a decision, but not to obtain an explanation of the decision reached. A right to explanation is only explicitly mentioned in Recital 71, but this not legally binding, and the authors examine how the development of GDPR excluded these provisions from the final text, moving them to the recitals.

The second legal avenue for a right to explanation would come from the notification duties of data controllers in Articles 13-14 and Recitals 60-62. These articles cover the information that must be provided to individuals before their data is processed. This includes the existence of automated decision-making, including profiling – pointing to Article 22 – and in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences. This seems more robust but in any case it would only apply before processing has taken place, when the data is collected, and before a decision has been made. This would mean that only general explanations can be given.

This right to information does not extend, in the view of the authors, beyond this initial point of collection. This argument is not clear, however, as there is an expectation that organisations will keep individuals up to date of substantive changes to their policies.

The third legal avenue for a right to explanation would derive from the right of access in Article 15 GDPR. The provisions are identical to Articles 13 and 14, but in this case allowing data subjects to request explicitly the information, as a mirror provision to the previous obligations to provide the same information. This article would appear to support explanations after the decision has been taken, given that a request can take place at any time, but the authors take the view that this would be inconsistent, as the phrase used in all the three articles, “envisaged consequences”, is future-oriented and therefore the right of access is not addressing how an individual decision was reached but simply an explanation of system functionality.

In any case, there would be some limitations to a ‘general’ right to explanation, applicable to all automated decisions, because the scope of GDPR is narrowed to only apply to decisions “solely based on automated processing” and with “legal” or “similarly significant” effects.

The word “solely” could create a loophole whereby any human involvement in a decision would exclude it from scope, although there are many argument for taking a broad view here. The Guidelines on Automated individual decision-making and Profiling by the Article 29 Working Party warns data controllers against trying to game these rules by inserting a nominal human in the process, who without taking other information into account would simply rubber stamp the decision.⁸

The authors admit that there are still good arguments in favour of the right to explanation. Having a right to contest a decision cannot be effective without understanding the basis for the decision, which would impact the rights to fair trial and effective remedy enshrined in Articles 6 and 13 of the European Convention on Human Rights.

⁸ http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 p.20

Unfortunately, GDPR protections against profiling and automated decision-making are more or less the same as in the Data Protection Directive 1995, and these provisions have not been sufficiently tested in courts since the original Directive came into force and these issues will require to be settled by new jurisprudence or in national legislation.

Machine learning and the transparency fallacy

In addition to the debate on the extent to which a right to explanation exists, there is a growing concern about the practical feasibility of implementing such right in the context of complex data processing such as big data, artificial intelligence and machine learning.

Lilian Edwards and Michael Veale⁹ argue that a right to an explanation is not the remedy to algorithmic harms. They support Wachter et al's doubts as to the legal basis for this right in GDPR, but even where explanation is provided in whatever narrow terms, as "meaningful information about the logic of processing", they argue that this is not compatible with how modern machine learning technologies are being developed.

The problems to tackle here are discrimination and fairness. Machine learning systems are designed to discriminate but some forms of discrimination are socially unacceptable and the systems need to be restrained. The general obligation of fairness in data protection provides the basis for the need to have some level of insight into the functioning of algorithms, particularly in profiling.

Their main departure is their proposal to partially decouple transparency as a necessary key step towards accountability and redress. They argue that people trying to tackle data protection issues have a desire for an action, not for an explanation. The actual value of an explanation will not be to relieve or redress the emotional or economic damage suffered, but to understand why something happened and helping ensure a mistake doesn't happen again.

Within this more limited sense problems remain to define transparency in the context of algorithmic accountability. For example, providing the source code of algorithms may not be sufficient and may create other problems in terms of privacy disclosures and the gaming of technical systems. They argue that an auditing approach could be more successful instead by looking at the external inputs and outputs of a decision process, rather than at the inner workings: "explaining black boxes without opening them".

The authors disagree with Wachter et al on the value of explanations under Article 15 GDPR, believing that it does provide some grounds for explanations about specific decisions. They present two types of algorithmic explanations that could be provided here: model-centric explanations (MCEs) and subject-centric explanations (SCEs), which seem broadly aligned with explanations about either systems or decisions.

⁹ Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for. SSRN Electronic Journal. <http://doi.org/10.2139/ssrn.2972855>

SCEs are seen as the best way to provide for some remedy, although with some severe constraints if the data is just too complex. Their proposal is to break down the full model and focus on particular issues through pedagogical explanations to a particular query, “which could be real or could be fictitious or exploratory”. These explanations will necessarily involve trade offs with accuracy to reduce complexity.

Their main concern seems to be to avoid creating a “transparency fallacy”, where similarly to the “consent fallacy” people get an illusion of control that does not exist, instead of being offered practical remedies to stop harmful data practices.

There is growing interest in explanation of technical decision making systems in the field of human-computer interaction design. Practitioners in this field criticise efforts to open the black box in terms of mathematically interpretable models as removed from cognitive science and the actual needs of people.¹⁰ Alternative approaches would be to allow users to explore the system’s behaviour freely through interactive explanations. This is quite similar to the proposals by Edwards and Veale.

A complementary approach has been put forward by Andrew Selbst and Solon Barocas,¹¹ who argue that the increasing calls for explainability of automated decision making systems rely on an intuitive approach that will not work with machine learning. ML is both inscrutable and non-intuitive. Inscrutability is the black box problem, the inability to understand the inner workings of a model, but non-intuitiveness means being unable to grasp the rules the model follows. Accountability requires not only knowledge of the process, but also whether it is justified, or fair.

Selbst and Barocas argue that lawyers and scholars asking for explanations will be disappointed because intuition cannot deal with the truly novel insights produced through machine learning that associate data in patterns that completely escape human logic and imagination.

Their alternative proposal is to focus accountability on the processes around ML models, not the models themselves. Policies and documentation of intent and design choices should be made available, some by default, such as impact assessment, and other in the context of a complaint or regulatory action. This approach chimes with the main thrust of GDPR’s accountability principle.

Right to portability

Article 20 of GDPR creates a right to data portability. This is a completely new right aimed at

¹⁰ Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. The 2018 CHI Conference, 1–18. <http://doi.org/10.1145/3173574.3174156>

¹¹ Selbst, A. D., & Barocas, S. (2018). The Intuitive Appeal of Explainable Machines. SSRN Electronic Journal. <http://doi.org/10.2139/ssrn.3126971>

giving consumers more control over their data. The right will allow users to move their data from one provider to another in order to improve choice and competition. This could mean obtaining the best tariff for a service based on previous usage, or making it easier to move your data around online services, preventing consumers being locked in to particular providers.

It is unclear how the new right to portability will work in practice. Consumers will have the right to freely download or receive certain types of personal data to make it easier to then send or upload the data to other companies without any obstacles from the original provider. This means that the data must be provided in a common structured format that can be easily processed by computers. This could involve formats for spreadsheets instead of text or pdfs. Where technically feasible, consumers should be able to get their data sent directly from one company to another. Companies should provide the data within one month, or maximum three if there are any problems, and must always respond to requests, even if to refuse them on any grounds. This may require extensive changes to the technical architecture of many companies and it is unlikely that most are ready to comply.

The right also includes being able to transmit the data to other companies, and this is an important element of achieving the objectives of the new rule. However, the obligations companies have to accept and process the data are not clear. The Regulation only says that data subjects have a right to transmit the data to another company.

There are some important limits on the types of data covered by this right. The right only covers data processed by automatic means, under consent, or where it is necessary for a contract. Only data provided by the consumer is covered. Data directly generated through consumption activities is generally seen as “provided”, but data generated by processing or analysing the basic data generated by the consumer – termed derived data – should in principle be excluded. Data created manually, for example a written assessment or scoring, will also be excluded.

The right to data portability overlaps with the right to access but it is not exactly the same, as portability covers a much narrower set of personal data and it is aimed squarely at promoting consumer choice, while access is there to help ensure that data is processed properly and enforce other rights. This is bound to cause widespread confusion.

The enjoyment of this right relies on companies making the data available in formats that are usable by the technical systems involved. While excel may be useful for consumers in a general setting, in some contexts portability may involve highly specialised data formats. The obligation covers every company and sector where data is processed automatically under consent or a contract. This is a large chunk of the consumer facing economy. It is to be expected that various sectors will develop standardised formats to be able to make this work, but this is not compulsory. European data protection bodies recommend that companies build tools such as application programme interfaces to enable direct data sharing .

Portability of data is already in place in several countries and sectors. The UK MiData has

provided some limited portability rights, for example in the energy and finance sectors, since 2012. This initiative has met with limited success as consumers were generally not aware of the possibility to obtain their data, in cases the data was not detailed enough – e.g. annual energy consumption instead of monthly – and there were limited tools to reuse the data.

Relation to other regulation

There are other rights to portability in various other legislation. One important area is financial services, where the new Payment Services Directive (PSD2) brings rights to portability to banking and finance. The UK has already mandated banks to give data access to third parties providing financial services using the OpenBanking standard, which is a very good example of the kind of regulatory intervention and industry collaboration required to make portability a success. Consumer and companies need to be clear of the framework under which the data is accessed, as this has practical implications.

Right to erasure / right to be forgotten

We are not going to discuss the right to erasure in detail, but it is worth pointing that the the pre-GDPR implementation of a right to be forgotten after the Costeja case at the CJEU¹² caused extreme controversy for its implications for freedom of expression and access to information.

The right to erasure or to be forgotten as implemented in GDPR seems to provide enough safeguards against excessive encroachment on freedom of expression¹³. However, request for erasure will continue to be controversial in certain contexts. Tools designed to exercise data rights will need to put extra care in their workflow design to avoid causing unwanted effects by enabling excessive unwarranted requests.

Requests for erasure of data that is publicly available, in registers or in certain sectors will need careful consideration. The now defunct Article 29 Working Party issued guidelines on the application of this right, that are now deprecated but still provide a useful starting point for dealing with such requests¹⁴.

¹² http://curia.europa.eu/juris/document/document_print.jsf?doclang=EN&docid=152065

¹³ Dulong de Rosnay, Mélanie and Guadamuz, Andrés, Memory Hole or Right to Delist? Implications of the Right to Be Forgotten for Web Archiving (June 1, 2017). 6 RESET Social Sciences Research on the Internet (2016). Available at SSRN: <https://ssrn.com/abstract=3107565>

¹⁴

https://web.archive.org/web/20150208095850/Guidelinehttp://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp225_en.pdf

Research Interviews

We conducted a limited qualitative analysis of the awareness and understanding held by UK citizens and consumers concerning their rights under GDPR. We did this in the period leading up to GDPR coming into force. Despite its limitations, we hope this exercise provide insight into public perceptions of data protection rights and help form a baseline to compare future similar analyses once that GDPR has been fully implemented. For each area, in addition to some preliminary conclusions we also provide some pointers for future research.

Methodology

We spoke to eight people for these interviews. The interviewees had responded to either a) a Twitter call from Open Rights Group's account or b) personal requests by members of Open Rights Group staff. We were particularly looking to attract people that did not consider themselves experts in data protection issues.

The interviews lasted for 20-30 minutes each and were carried out over the phone. They were semi-structured in that the same themes were covered in each interview, but the order and phrasing of the questions varied. This helped to keep the conversation as natural as possible and elicited useful information.

We informed interviewees of the nature of the research and that we planned to write up what they said for this report. We obtained their consent to make an audio recording of the interviews and to write them up with names changed and identifying details removed.

Research Themes

We wanted to gain a better view of how people who are non-experts think about data protection issues. As a top level question, we wanted to see how people understood data protection law and their data protection rights. We also wanted to see how they thought organisations should behave when collecting and processing data about them.

We covered topics such as whether they saw a hierarchy of sensitivity of data. Did they think that some data was more sensitive than others and should use of those data be more tightly regulated as a result?

We also talked to the interviewees about their sense of power and control over data about them. How did people experience the emails that many organisations sent out due to GDPR? Did they feel informed about the way their data was being used and did they feel in control?

Another area we looked at was how people would do if they wanted to raise a complaint about a data protection issue. What avenues did people expect to be open to them? What

has their experience been if they have previously raised a complaint about a data protection issue?

We also found that some themes emerged about good ways in which to communicate data protection rights from the perspective of Open Rights Group, the Information Commissioner's Office, and from organisation that collect and process data.

We wanted to get an understanding of how people would engage their various data protection rights to erasure, access, rectification and so on. As we explain later, we could not acquire enough data to discuss people's understanding in detail. This is a useful finding in itself.

Research Findings

Data categories and their relative sensitivity

One interviewee worked in a job which meant she wanted to avoid lots of people being able to find out information about her online.

"I would prefer if you Googled about me and absolutely nothing came up. I would prefer to be anonymous."

She said that she 'de-tags' images of her "religiously" out of a fear that her image could be abused without her knowledge:

"Could somebody out there be collecting images of me and could my image be being used in a way I wouldn't want it being used?"

Another person we spoke to had few qualms about companies collecting her shopping preferences in order to serve her with tailored advertising online and in the post.

"If I go on Google and search for something, say a pair of shoes, then those shoes will pop up all the time. So that has registered with Google that I'm looking for that item and they're suggesting other similar things for me. I don't actually think too much about it. It just seems to be part of our lives now that this information is held, stored, and used. I think the law says if it's used for its written and intended purpose which I guess if I am looking for shoes then it is its intended purpose.

...

If I for example request a clothing pamphlet for a kind of clothing then I find I get sent other similar pamphlets and I haven't requested those so they must be giving my information to other people, to other companies and they think I might be interested in other clothing too."

She added that she thought of this sort of marketing as a company's "right":

“I don’t actually mind too much cause I think well that’s their right actually. These are the kind of clothes that I might wear and it’s not really doing any harm I just throw it in the bin.... Sometimes it works, sometimes it doesn’t.”

An interviewee who had worked for a direct marketing company said that she never mentions her name or location on social media and never posts photos of herself online. When asked whether her experience of seeing how data about people can be used affected how cautious she is with information about her online she answered:

“Definitely. Actually being aware of what you can do with data drives part of my caution about how I let mine be shared.”

One interviewee talked about how although it would be easier to market to him using data about his hobbies and how he spends his free time, he is far more concerned about organisations and government collecting his political views.

Similarly, during a conversation about political views being used for profiling, another person we spoke to talked about how he disliked the idea that organisations would collect or infer his political leanings and then use that data about him:

“I wouldn’t like to think I’m going through life having decisions made about me because of things like that [his political views] and any kind of leanings or views at all.

...

On social media, lots of the information about me is inferred rather than given so I suppose I dislike the idea that knowledge about me would be extrapolated in some way rather than freely given. If Facebook have decided that I’m a Labour voter, it’s not because I’ve told them. It’s because they’ve done analysis, automatically or otherwise.

...

It’s partly because it’s inaccurate and partly because it’s none of their business.”

Conclusions and future research

Broadly the data that the people we spoke to felt was sensitive, seemed to align with the categories of data that GDPR specifies as sensitive data. In other words, their instinctive understanding of the law lines up reasonably closely with the actual law. Political leanings came out strongly as something a couple of our interviewees thought was particularly sensitive. (This may be somewhat affected by our sample of people coming from replies to a tweet from a political campaign group, of course.)

The interviewees felt the sensitivity of data was in some cases highly influenced by the context in which the data could be used. The interviewee who was concerned that no-one related to her professional life find out about her personal life was particularly conscious about the sensitivity of photos of her.

While most of the people we spoke to were fairly protective of the information about them, one of our interviewees was quite relaxed about companies knowing her browsing and search history – the items she was looking to buy online. She saw this as part of living in today's world and not especially harmful as you could just unsubscribe from marketing. In future research, it would be interesting to explore further how context affects how sensitive people think certain categories of data are. How common is it for political views and visual likeness to be considered sensitive, for example?

How organisations should behave when collecting and using data

Several of the people we spoke to put a very strong emphasis on consent when asked what companies should do if they want to collect and use data about them. This was one example:

“Well I just think it’s explicit agreement to say from our side [companies] ‘This is what we’re going to do XY and Z with your data.’ and from your side [individuals] ‘Are you happy with this? If yes, great. If no, then we won’t do XY and Z with your data.’ And I think something as explicit as that would have been a good idea.”

The same interviewee later talked about companies that were refreshing consent due to GDPR and drew a link between consent and political correctness:

“Companies in the social network and service sector, their political correctness was good.”

Another interviewee linked consent with politeness and transparency in her reasoning for why consent was important.

“I feel like I should be consulted if that’s going to happen, obviously if it’s personal information. But even if it’s just your name and the few things that you put up on your Facebook profile, I feel like you should be consulted. I don’t feel like that should just be open to the world.”

...

“If I was going to an interview, I feel like I should be asked if they want to Google me. That’s polite and open and transparent. Trying to find out about somebody and doing it without their knowledge, it feels a bit sneaky.”

In a similar vein, another interviewee said:

“I don’t like that I have no say over they way they can use my data.”

...

The way that Facebook have been shown to use people’s data, I think that’s unacceptable. Nobody consented to that.”

One person we spoke to put a much greater emphasis on regulating what organisations can do with data than on what they should do to collect it.

“I’d like there to be much stricter rules about who they can sell it [data about him] to and how that can be applied. The notion that not just my details but the details of people I’m in contact with can all be gathered up to generate a map of where I live and what I likely believe and what I’m likely to be swayed on and that the fuel that makes that possible can be taken from me without my knowing just cause I wanted to play Farmville or something, I’d like that to be punishable.

...

The idea that data about people can be traded in the same way that a corporation’s assets can be traded, it’s not the same as a printing press that you’d have to move from the factory - it’s not a cash asset that’s degradable in some way. The information about all of us in aggregate - that’s hugely powerful and valuable. And people it’s been drawn from, they get neither profit from it or awareness that it’s being done. It just seems grossly unfair”

One of our interviewees talked about the importance of allowing organisations to store data even without consent:

“I think it’s important that there are restrictions on what data can be kept. But at the same time I also think it’s important that the data can be kept where there is a genuine need for that data to be kept. Otherwise you get reciprocal problems where the data is being discarded when it actually needs to be kept. A good example of that from our work is that we sometimes see people who would perhaps haven’t filled out the paperwork properly for consent but in fact their situation is very bad and then we have to make a decision for ourselves about whether we need to keep that data. And on occasion we do because we look at it and we say we can’t afford to lose these people’s records because that will put them in danger or make it harder for them to access services.”

Conclusions and future research

There was a fairly uniform belief that organisations should obtain consent if they want to collect and process data about people. Awareness of the other lawful bases for processing data was very low, sometimes even among people who had a well-developed understanding of data protection issues in general. Some interviewees insisted that consent was necessary even when prompted to consider whether banks and employers should always need to gain consent to process data about them. In several cases then, people’s assumptions about what the law should be do not align with the law itself.

Although we have a small sample size, it was remarkable that the only person to talk about how it was sometimes important to process personal data without consent had direct professional experience of the issues involved.

It was particularly interesting to hear how tightly some people linked consent with concepts like transparency, decency, and political correctness. One area that would benefit from further exploration would be what positive concepts can be used to link with the other lawful bases. It would also be useful to find out whether people stress consent so strongly when there is more time since they have received lots of emails asking for their consent. It may be that people talked about consent because it was still at the forefront of their minds thanks to GDPR consent refresh emails.

Control over data and GDPR emails

People had very varied experiences of emails from organisations about GDPR to either refresh consent or inform users about new privacy policies.

One interviewee seemed to positively relish the experience and actively unsubscribed from marketing databases:

“I found it a brilliant opportunity to unsubscribe from lots of databases that I didn’t realise I was still on. I did have a few moments of surprise where I remember thinking, well I certainly don’t remember signing up to anything related to this company so where have they got my details from. They were mostly recruitment companies.”

One of the people we spoke to was also positive about the experience but was happy just to let subscriptions lapse. They reported much lower numbers of emails in their inbox since GDPR came in:

“I just completely ignored all of them because I thought, if it matters then I can always re-subscribe to this thing. But actually it’s been so amazing since May the 24th was it that my inbox is just so much more manageable.

...

Most of them I knew why they had my email but I just hadn’t got round to unsubscribing from them. There were one or two things like courses where I’d just completely forgotten they existed. Now I feel like it’s just revolutionised how much time I spend on my phone for emails. Yeah I am starting to feel more in control.”

Another person we interviewed thought it was a good thing they were being better informed about how to make inquiries to a company about data protection but was frustrated that he was receiving emails that perhaps were unnecessary:

“I took it as an opportunity to remind myself of who I was signed up to. Some of it was just annoying. I don’t really believe they needed to contact me at all but it did remind me that they have my details.

...

I did sense that companies were taking it [data protection] more seriously. I also feel like you’d be more responsive. Now they’ve got data protection officers they do tend to have a mailbox dedicated to privacy issues and they will respond to you. So I do

feel like even if I didn't have more control, I do have more of an avenue to make contact and make inquiries. Yes it does make you feel like you have a bit more of a say in things."

Others displayed scepticism about the long-term impact of GDPR. One interviewee thought that companies may well revert to sending marketing messages in a few month's time:

"I suppose it was comforting to believe that there were going to be some standards around how to deal with this and whether they deal with data properly and use it in some way that they could be held to account for. Long term it would be nice if you could think that the principles behind it were going to be upheld. Give it about two months. We'll see if anybody's still bothering. I'm sure that the organisation that's supposed to be chasing up compliance will find some people who fall foul of legislation and make a public example of them. But by and large I expect that we'll go back to business as usual in a couple of weeks."

Another person we spoke to did not feel a sense of control about data about them. They thought that companies were going through the motions to some extent and doing their "duty":

"No I didn't really. It felt like an independent process that was put in place for good reason and was grounded in a sensible view of standardising across the European Union procedures about access to data and information. But equally I felt like it slipped into a formality of doing one's duty."

Conclusions and future research

No two people we spoke to had the same experience of GDPR emails or of the sense of control they had over their data thanks to GDPR. We did not find so-called 'consent fatigue' to be particularly prevalent among the people we spoke to.

Supporters of GDPR including some politicians have talked about GDPR as a opportunity for people to have greater control about data about them. It may be that the legal rights allow individuals greater control over data about them. It is very unclear however whether this is the actual experience of the people we spoke to. Trust in the effectiveness of law was low in one case. Elsewhere, we heard people talking about how they still did not feel truly In control of their data despite GDPR.

It was also evident that emails from companies to refresh consent or inform people about updated privacy policies was overwhelmingly the main way people had come into contact with GDPR. Awareness of rights to portability, to erasure, and around automated decisions was very low. Further work will need to be done if these rights are going to be used and understood widely. Another interesting avenue for future research would be to investigate in greater detail the other ways in which people can feel a sense of control over data about them. This would likely link closely to a greater awareness of the GDPR rights and how to engage them.

Communicating about data protection and GDPR rights

Some people talked about how they thought data protection issues in general as well as GDPR rights more specifically should be communicated.

One person thought that much greater clarity from business about what data they collect and how they use it would be beneficial for both individuals and the businesses. He thought that unclear and long privacy information was causing some people to opt out and some people to opt in without thinking:

‘What’s missing at the moment is that there’s not enough plain speaking. So you’ve got either people who’re looking at it and going I’m just going to click ok cause I just want to move forward and you’ve got people who read it and go actually I haven’t got time to read all of this, I’m just going to opt out of it. Or who read it and go I don’t understand so I’m going to err on the side of caution. And I think from a business standpoint that’s going to hinder them. I think using simpler up-front transparent language would be better. If you say as a business, this is what we do as a business, here’s how we’re going to use it and here are the limitations we’re going to put, you would probably get more engagement as a business.’

Another interviewee, when asked about the things that would help her friends and family care about their data like she does, answered that good stories were the way to persuade people:

“Actual conversations over dinner table. Normally the way of getting through to people is stories. This is a thing that actually happened. People respond to narratives that they can relate to and go, wait what about that? The woman who in the US actually got sent baby-related material because she was pregnant before she knew she was pregnant. That may be a complete coincidence but it makes a good story to communicate this to people”

She added that the Cambridge Analytica story was also a good way to help people understand the value of data protection:

“Obviously, all the stories around Cambridge Analytica in the last month have turned around doubters who would have fought you tooth and nail and now you can say no people are up to things that you don’t want to think about.

...

People don’t like being confronted with the idea that maybe their actions weren’t entirely what they would have been. They don’t like knowing they’ve been manipulated.”

Another person, completely unprompted, started talking about Cambridge Analytica and the effect of the story on their view of data protection:

“I think recently the Cambridge Analytica thing has just helped everybody to feel a little more aware of websites and the power that they have. And I feel like since that scandal came out people are getting the power back a little bit more over organisations and I think that’s a really good thing.”

Conclusions and future research

Although we did start out looking for our interviewees to offer ways to communicate about data protection rights, some of them brought up the subject. It was interesting how the interviewee who talked about the need for greater clarity of communication from companies about the data they collect and their purposes framed it as being of benefit to companies as well as individuals. It would be valuable to look again at how the elements of GDPR around transparency and intelligibility can be communicated to businesses as being of value to them.

There is something about the Cambridge Analytica story which has really grabbed people. Many of our interviewees mentioned it in passing and some delved into what they thought the effects of it have been. It feels like the interviewee who talked about how people don’t like being manipulated and the corresponding sense of lack of control has identified the key.

It would be worth continuing to explore how the GDPR rights can be communicated in such a way as to build trust that the rights do allow personal control over data. It seems unlikely that mandating their inclusion in privacy policies is the best way to communicate these rights. Social media and press campaigns may prove more likely to improve awareness and trust in the rights.

Complaining about data protection issues

Some people had not heard of the Information Commissioner’s Office. After being asked who he would complain to if he had a problem relating to data about him, one interviewee replied:

“That’s the problem. Maybe the fellas down at the pub? Who do I think could actually make a difference? Some well-meaning journalist? There is no redress.”

Another interviewee who was generally very well informed about data protection issues appeared to be unaware of the existence of the Information Commissioner’s Office. He said that if he had an issue around data protection he would contact the company in question but said that if he wasn’t happy with the response there is no regulating authority:

“The only other recourse would be the courts.”

One interviewee said he would contact the Data Protection Officer of an organisation first. Then if he felt they were being difficult he would contact the Information Commissioner’s Office. He had experience of complaining to the ICO and was sceptical that complaining to the ICO would come to much in most cases:

“It feels like throwing stuff into a black hole. I don’t doubt at all that they do good work. But I think if you’re an isolated case or you’re one of a handful of people who’ve got a grievance with an organisation I just don’t think they’ve got the resources to deal with that. I don’t really believe that those sorts of complaints are going to come to anything. It’s the big companies, that’s the level that they work at. So if your local hairdresser is putting their customer contact details in their Dropbox account and breaking the law, nothing’s ever going to happen about it. If that information goes walkabout then it’ll just be Dropbox that get in trouble and not the hairdresser. With small scale stuff I just don’t think anything’s going to change.”

Another person we spoke to was sceptical about the utility of complaining about data protection issues, including to the Information Commissioner’s Office.

“Quite honestly I don’t know where you’d complain. I don’t see the benefit of complaining corporately. I’ve not been given any indication that that’s practical. I think if there’s a breach I might complain to the Information Commissioner. But if I’m honest I don’t see any value in that. I don’t feel that that organisation has got the strength needed.”

Conclusions and future research

It seems clear that awareness of the complaints system for data protection issues is low among the people we spoke to. Some people were unaware of routes through which they could lodge complaints even when they had relatively high levels of understanding of data protection issues in general. Levels of confidence that they could be successful in having a complaint resolved to their satisfaction were also low.

Several people were aware of the Information Commissioner’s Office but said they wanted the ICO to have and use greater powers and resources to deal with smaller scale infractions. The only case that any of the interviewees volunteered was that of Cambridge Analytica. This suggests that awareness of investigations into cases such as that of the University of Greenwich or the Bible Society was low.

Clearly for people to use their GDPR rights to complain to a regulating authority they have to be aware of the relevant regulating authority. When they are aware of the regulating authority, in most cases they also need to be relatively confident that their complaint has a good chance of coming to a satisfactory conclusion if they are to use their right. Conclusions and future research from our interviews would benefit from further study. They appear to suggest however that some work should be done to build up awareness and confidence in the GDPR rights around making complaints.

Engaging data protection rights

One of the areas we were interested in exploring in these interviews was how widely the various GDPR rights to erasure, rectification, access and so on were understood. None of the people we spoke to talked at any great length about how to use these rights or why they

would use them. It may be that the questions we asked did not elicit this specific information as well as they could have done. A more tightly structured interview where we asked scripted questions to the interviewees could have been more successful at bringing their understanding of the GDPR rights to the fore.

On the other hand, the people we spoke to brought up lots of topics in the interviews that, while relevant to the conversation and our research, our semi-structured questions were not necessarily designed to cover. In other words, it seems likely that the people we spoke to brought up the topics about data protection that they were most conscious and aware of. This may mean that the rights to erasure, rectification, access and so on are not among the data protection topics they are particularly familiar with.

Conclusions and future research

Data Rights Finder¹⁵ – the site we have made as part of this project – was planned with the premise that people need help to understand what their GDPR rights are and how they can use them. We will test this proposition over the next few months as people use the website. That none of our interviewees brought up these specific rights or recognised them when prompted may suggest that our premise is broadly correct.

Future research could usefully focus precisely on public awareness of the various GDPR rights. It would be interesting to understand which of the rights are most widely known about. For example, are people more likely to know that you can get a copy of your data from an organisation compared to that you can review an automated decision about you? It would also be worthwhile to understand how valuable people think the rights are once they have the rights explained to them.

¹⁵ <https://www.datarightsfinder.org>

Vox pops about Data Rights Finder by Project by IF

Separately to the research interviews discussed above, Projects by IF carried out vox pops around a very early prototype of the tool that became Data Rights Finder.

They approached four random people around Somerset House in London, and interviewed them for 10 minutes each. The four people gave their consent for us to anonymously use quotes from these interviews. IF showed them two wireframes that illustrated a web service where people can search for an organisation and see information about how to exercise their GDPR rights.

To start the interview, IF asked participants whether they were aware of GDPR, and if so, to explain what it is. IF then showed the first wireframe and asked participants:

- What do you understand about this wireframe?
- Would you use a service like this?
- If so, when would you use it?
- If not, why not?

This is the first wireframe.

○ ○ ○

https://rightsfinder.projectsbyif.com

RightsFinder

Enter a company name... **Search**

About RightsFinder

The General Data Protection Regulation gives people new rights to see and control how data about them is used by companies.

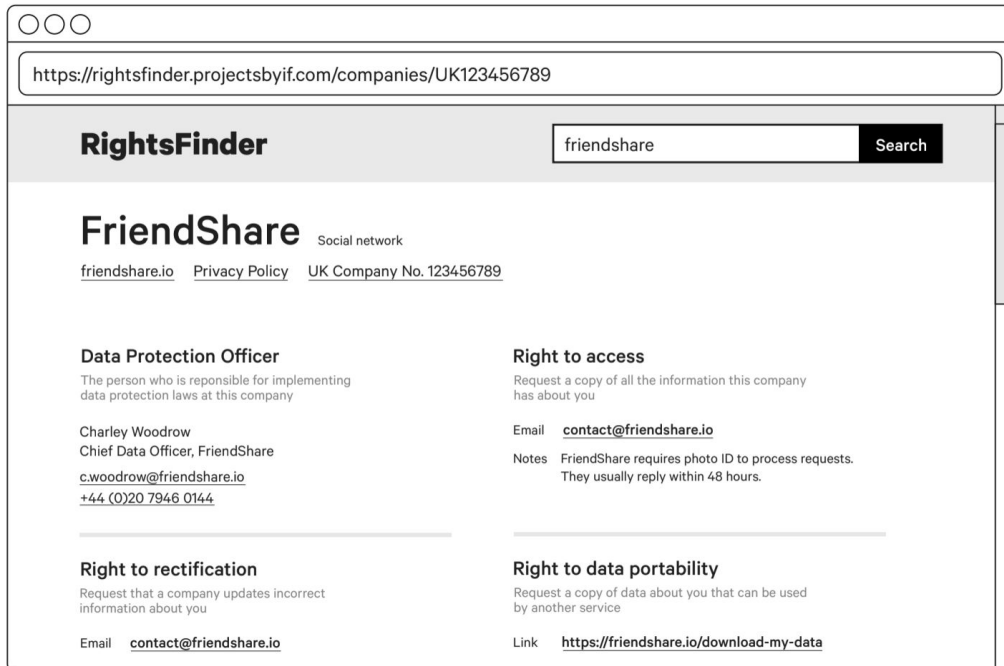
RightsFinder collects the information people need to use these rights, helping them find the links and contact details that are often hidden in hard to read privacy policies.

IF then showed the second wireframe and asked:

- What do you understand about this wireframe?
- What information would you find useful?
- What information would you find not useful?

- What information do you think is missing?

This is the second wireframe.



Findings

Most people knew what GDPR is. People were aware of GDPR through their work, but not everyone showed confidence in their understanding. One person said "I'm familiar, but not as much as I should be". Another person "understood the basics" but relied on their legal team. One person they said they did not know what GDPR was.

People responded well to having access to summarised and actionable information. A participant said that the prototype showed a "helpful summary". Another said, "If I wanted to get off a mailing list, I would know who to contact". Actionability of this information was also highlighted. One participant said it is "useful to have clickable links".

People suggested this service would be useful in a professional context. A participant said they are "a bit laissez-faire about personal data, but I might use it check for work". Another said they would "Probably not [use it for] personal use, more professional".

People want to see more than just raw information While this service was propositioned as a lookup tool, participants wanted supporting information, particularly around the new GDPR rights: "Maybe more clarification about what those rights mean." The avoidance of not showing an entire data set at once was also highlighted: "It's useful but could be ordered more logically".

Survey of data rights tools

This section provides an overview of some of the tools currently available, or in development, that aim to support data rights. There is growing interest in technical means to improve data protection and privacy, of which Data Rights Finder is part. We describe the tools in simple terms and we are not providing an in depth analysis of each. In most cases the information comes from simple tests of the websites, although in others this has been complemented with interviews with the developers, particularly in two of the tools that have not been launched yet.

The tools we reviewed are not exclusively from the European Union. We have not covered any tools aimed at supporting GDPR compliance for companies. There is currently an explosion of online tools and services to help companies prepare and document their GDPR compliance activities. Looking at how well these tools deal with user rights would be useful, but it is out of scope. Besides, it is to be expected that this market will consolidate and many of these tools will disappear or be bought by larger vendors.

The tools we reviewed show a broad range of approaches to privacy rights. There are two main categories of tools here, those that aim to help an individual exercise specific rights and those that mainly aim to improve access to company information, mainly privacy policies.

All the tools for rights support individuals with requests to access data, but they vary on their approach, from the very lightweight generation of templates and pointers to an email address or web form, to sophisticated systems for managing ongoing correspondence with organisations. In almost all cases, the responses are sent directly to the individuals, and we perceive a reluctance to handle the personal data of their users, understandably. Some of the tools expand to enable other rights such as erasure, correction or even portability. The latter is particularly problematic, and sector specific, and we still have not seen a good general solution.

We showcase two tools to facilitate the understanding of privacy policies, which take completely different approaches. Polisis is a machine learning automated system and ToS;DR is based on manual analysis. Both use innovative visual techniques to facilitate the display of information to the users.

Interestingly, none of the tools aim directly to simplify and structure the way that companies write and display privacy information themselves. We understand that there are other projects not covered in this report working on data structures for this purpose¹⁶. Engaging with companies will be an important element in the development of these tools, particularly sympathetic companies that genuinely wish to improve the way they handle requests from their customers.

¹⁶ The venerable P3P standard from W3C attempted to implement a common machine-readable standard for websites but it failed due to the complexity of the field, and since then any similar efforts have been generally received negatively among internet technologists. <https://www.w3.org/P3P/>

One common theme to the projects is the effort required to input the information about companies into the database. The automated approach of Pribot is promising but it seems to require extra human support. We believe that the data structures and backend developed in Data Rights Finder can be a useful contribution to this field.

PersonalData.io

<https://personaldata.io>

The PersonalData.io project is a tool to facilitate subject access requests. It is run by Swiss mathematician Paul Olivier Dehaye, who has gained notoriety through his collaboration with the Guardian journalists investigating Cambridge Analytica. This led to him giving evidence to a committee in the UK House of Commons. Dehaye has been using subject access requests for personally driven privacy campaigning for some time now, as documented in his blog¹⁷. The tool is currently free but Dehaye had expressed in the past that he was planning a commercial service.¹⁸

Campaigns

What could you do if only you had the data? Don't hesitate to contact us if you have a project that requires the collection of personal data. We provide the necessary expertise and infrastructure to groups of people for data-collection campaigns to be conducted efficiently at scale so you can focus on the other aspects of your project.

The screenshot displays three campaign cards on the PersonalData.io website. Each card features a logo at the top, a title, a short description, and a 'Read More' link with a heart icon and a count.

- TINDER:** The logo is in red. The text reads: "TINDER. Tinder predicts a desirability score, which it uses to match you with others 'in your league'. Do you really want to find it out???" Below the text is a 'Read More' link and a heart icon with the number 132.
- UBER:** The logo is a green square with a white 'U' and the word 'UBER' below it. The text reads: "UBER. Curious of how Uber tracks you? During and outside of a ride? Through your phone or the drivers?" Below the text is a 'Read More' link and a heart icon with the number 45.
- FACEBOOK EMOTION MANIPULATION EXPERIMENT:** The logo is a blue square with a white 'f' and a hand holding a scale. The text reads: "FACEBOOK EMOTION MANIPULATION EXPERIMENT. Were you part of Facebook's controversial 2012 experiment on mood manipulation? How did Facebook try to affect you?" Below the text is a 'Read More' link and a heart icon with the number 39.

At the bottom of the cards, there is a blue button labeled "All campaigns" and a grey button with a mail icon and an upward arrow.

The tool has a listing of under 30 companies that users can request data from. There is no apparent common criteria, and those may simply be organisations that the developers have engaged in the past for their campaigns. The organisations range from the well know - Twitter, Spotify – to the fairly obscure, e.g. Lingvist or BNI. Surprisingly, there is no entry for Facebook, despite the developers having engaged extensively in sophisticated requests with this company. There is no further information about the companies, not even a URL or contact details. In some cases this makes it difficult to know which company we are dealing with. For example, there are two Spotify companies and a Google search from BNI seems to point towards Business Network International, but this is not 100% clear.

Users need to register with the site and are asked to provide further details for each company. Once they click through they get a banner saying that their submission has been

¹⁷ <https://medium.com/@pdehaye>

¹⁸ <https://2040infolawblog.com/2018/05/27/masterclass-in-not-answering-questions/>

received and they will send the request shortly. There is a simple dashboard listing requests submitted, but no way to keep track of responses or clear information on how the response will be received. The privacy policy says that on rare occasions responses may be “routed through” them.

In addition to requests to companies, the tool has a “campaigns” section, that at present has one running on the Online advertising ecosystem, looking at internet tracking and advertising on mobile devices. Users need to install a third party app on their devices to obtain the unique identifiers used in the requests.









My Data Request

<https://mydatarequest.com>

My Data Request is a slick website that facilitates requests for data access to over one hundred companies. It covers a variety of mainstream online services - Facebook, Twitter - games, dating sites, airlines, and other assorted businesses.

Where do you want to request your data from?

Search for a company, e.g. Facebook

 Tinder	 Facebook	 Uber	 Lyft
 DoorDash	 GrubHub	 LinkedIn	 Slack

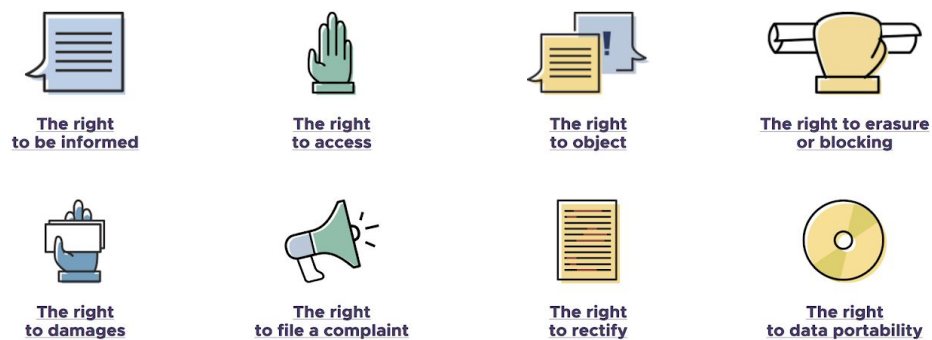
For each company the site displays very limited information, just a link to the privacy policy, contact information and a template letter that can be copied. There is no facility to send requests directly. The letters are customised depending on whether the user is based in the EU or elsewhere, with some special options for California residents in some cases. For Google, the tool points to the company’s data download service - Takeout, although it is unclear this constitutes a proper answer to a full subject access request.

The tool is basic but very well designed and the developers seem keen to engage their users, even offering a service to help with requests, and also for companies to help them improve their processes. The main issue is that there is no information as to who is behind this website. The email address provided is generic for users of Amazon Web Services hosting.

Philippines: Know Your Data Privacy Rights

<https://privacy.gov.ph/know-your-rights>

This website from the Philippines National Privacy Commission, the data protection authority, is an educational tool for end users. The Philippines Data Privacy Act (RA 10173) was enacted in 2012 but secondary legislation required to implemented was delayed until 2016. The majority of the provisions are based on the EU Data Protection Directive, and it is closely aligned with the GDPR¹⁹.



The tool is a very good example of institutional education, but it does not provide any templates or contacts for companies. Given that many organisations will be registered with some data protection authorities and will have contact details, it would be interesting to explore what an institutional tool for enabling rights may look like.

Open Schufa / Selbstauskunft

<https://www.openschufa.de>

<https://selbstauskunft.net/schufa>

Open SCHUFA is a joint project of Open Knowledge Foundation Germany (OKF) and AlgorithmWatch (AW). The main aim of the project is to crowdsource insights on the inner workings of the German credit rating system – SCHUFA – by getting volunteer to “donate” their credit reports to the platform. The SCHUFA system has financial information on most German residents, with some 10% of them having negative credit markers. After a one-month crowdfunding campaign in February / March 2018, around 1,200 supporters

¹⁹

<https://iapp.org/news/a/gdpr-matchup-the-philippines-data-privacy-act-and-its-implementing-rules-and-regulations/>


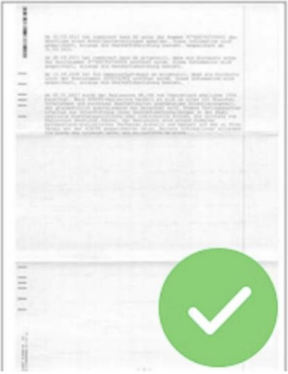
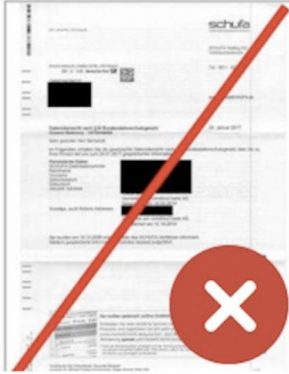
gathered nearly 44,000 euros for the development of the open source software that powers the main site²⁰ and the implementation of the project.

Los geht's!

Damit wir die SCHUFA knacken können, brauchen wir von Dir die Fotos oder Scans Deiner SCHUFA-Daten. Anschließend bitten wir Dich um einige Zusatzinfos von Dir.

Welche Seiten wir benötigen

Wir benötigen von Dir zwei oder ggf. mehr Seiten der SCHUFA-Auskunft, die eine Tabelle und weitere Infos enthalten. Seiten, die oben recht das SCHUFA-Logo zeigen, benötigen wir NICHT.

Nächster Schritt

As part of the process, volunteers request their information using a separate platform called *Selbstauskunft*, which translates as self-assessment. This request platform is run by a private company and has been operational since 2010 and has been used over a million times. The platform includes an impressive directory of 1839 companies including credit bureaus, banks, utilities and other private sector companies plus 1383 public sector “registration authorities”. For each company the tool contains basic information and contact details.

Users of the *Selbstauskunft* tool can make requests for personal information to an unlimited number of organisations simultaneously, but the responses will normally be mailed by post directly and not received via the website. In some cases the requests require a copy of ID documents and the platform allows the hosting and sending of these. The requests for self-assessment information on financial status is used in many situations in Germany, such as when renting accommodation.

²⁰ <https://github.com/algorithmwatch/openschufa>

My inquiries

Filter: Open requests ✕

UNIVERSUM Business GmbH (formerly Producta Daten-Service GmbH) 1806 KHW2CF JSON XML

06/28/2018 Request for self-assessment PDF

✓ even shipped on 28.06.2018

answer received Complete the request Create a reminder take care of

SCHUFA Holding AG 1806 XCSYZ3 JSON XML

06/28/2018 Request for self-assessment PDF

✓ even shipped on 28.06.2018

answer received Complete the request Create a reminder take care of

infoscore Consumer Data GmbH 1806 R9BNF7 JSON XML

06/28/2018 Request for self-assessment PDF

✓ even shipped on 28.06.2018

answer received Complete the request Create a reminder take care of

Here you can filter your requests by status and year.

REQUEST TYPES

Request for self-assessment 7

OPEN REQUESTS

withheld 4

shipped 3

The tool generates customised letters in the PDF format, and will send them directly to the companies. There are options for the user for printing and sending the letters directly. The tool has mechanisms to keep track of responses, including setting reminders. The data on the timing of responses is used to generate charts.

The OpenSchufa project leverages this well crafted platform to enable users to obtain their SCHUFA credit reports and upload them to their website. They have received over 10,000 requests, with double that number to other scoring agencies operating in Germany. To date they have not published the results.

The OpenSchufa project raises some very important questions about the coordinated use of access requests for – although in Germany access to credit reports appears to be regulated independently of GDPR. The use of requests for research based on collaborative data collection and to improve algorithmic transparency is very promising, but as we discussed elsewhere in this report, it is possible that there will be some resistance. Reverse engineering algorithms through coordinated requests for data may well be in breach of intellectual property legislation.

My Data Done Right – Knowing what’s known about you

My Data Done Right is a project by the Dutch digital rights organisation Bits of Freedom²¹. It has not been publicly launched yet, and the information below – provided by the developers – may be subject to changes.

The web based tool aims to help its users exercise their rights under the new General Data Protection Regulation (GDPR). The tool focuses on four specific rights: enabling access

²¹ <https://www.bof.nl/>

requests, correction of data, remove of data and moving data under portability rights. The tool is designed in the form of a multi-step questionnaire guiding users through the process. The responses will be sent directly to the users and not to the platform, which will not hold any of the data. The tool will have an optional way for people to get a reminder to send a follow up letter if they don't get a response, which we will also be provided.

The platform will ask people whether they would be willing to give feedback about the response. This will consist of simple questions such as: "Did you get a response in time?" or "Do you feel that data that your received was complete?"

The tool will start covering companies in the Netherlands, but aims to be EU-wide in scope at later date, and not limited to a particular industry or sector. The data will be crowdsourced by volunteers. Users will be able to request the addition of particular companies to the tool.

Access My Info


<https://accessmyinfo.org>

Access My Info is a tool run by The Citizen Lab at the University of Toronto in Canada. It aims to help people to carry out Data Access Requests – roughly equivalent to Subject Access Requests in the UK – with three categories of companies: dating, fitness trackers, and telecommunications. The site generates a letter that can be posted or emailed to a company's privacy officer. It is available in both English and Canadian French.

Welcome to Access My Info!


What do companies know about you? What do they keep on file? Who do they share it with? Organizations are required by Canadian privacy laws to disclose this information to their customers upon request. We can help with that.

Request information from:




Dating Applications

Your personality traits, sexual preferences, dating history, and other lifestyle information.



Fitness Trackers

Your heartbeat, sleeping patterns, diet, weight, walking habits, and general health.



Telecommunications

Your phone call records, web browsing history, geolocation, and device identifiers.

You may make multiple requests with our website (but one at a time!).

The user selects which category of company and then which company they want to contact. They can then select which categories of data they want to ask for. This is presented as a ticked-by-default checklist of categories: IP address logs, subscriber information, geolocation data, mobile app data, any other data from mobile/web services, personally identifying information, and lifestyle information. By default the user would request all of these categories and would have to explicitly deselect a category if they did not want to request that data. They then enter their name, email address, phone number and username so the company can identify their records and the site generates a letter which they can email or print off as a PDF.

The site offers information about why the right to access is important, how users can expect companies to respond, and what sort of data they can expect to receive.

A disclaimer in very smaller text in the footer of the website says that the site is a research and education tool that is meant for informational purposes only and does not provide legal advice. All of the content on the website is licenced under a Creative Commons Attribution-NonCommercial-ShareAlike licence. This means that the data on the site can be reused. This is commonly considered the most restrictive Creative Commons licence however so the potential to reuse the data is relatively limited. The Citizen Lab's main website is licensed as Creative Commons Attribution suggesting a conscious decision was made about the licence on Access My Info.

The Citizen Lab released a report in February 2018²² which analysed the barriers to access requests and the data that companies were releasing in access requests. It used data collected through the Access My Info website to write the report.

DataRights.me

<https://datarights.me>

The DataRights.me tool is currently under development and will be launched in July 2018. The information we have comes mainly from interviews and may change when the tool goes live. DataRights.me is developed by a group academics at TU Delft in the and VU Brussels²³, as a collaboration between participants and academics to collate research data on company behaviour around data protection and user rights. In the future their aim is to be able to research practices of organisations based on the responses and check differences from their privacy policies.

The open source tool²⁴ enables subject access requests, but with a focus on the follow up process and obtaining feedback for research. For example after a month the tool can send a reminder to the user to contact the company to remind them of the deadline, or a data protection authority to put a complaint. Users can send the requests directly from the tool but the replies

²² <https://citizenlab.ca/2018/02/approaching-access-look-consumer-personal-data-requests-canada/>

²³ Mahieu, Rene and Asghari, Hadi and van Eeten, Michel, Collectively Exercising the Right of Access: Individual Effort, Societal Effect (December 1, 2017). GigaNet (Global Internet Governance Academic Network) Annual Symposium 2017. Available at SSRN: <https://ssrn.com/abstract=3107292> or <http://dx.doi.org/10.2139/ssrn.3107292>

²⁴ <https://github.com/datarights/DataInSight>

will go directly to the user, although they are building facilities for the users to make it easier to share the content of the responses with researchers after receiving them, if they choose to do so. The developers are not planning to build facilities for bulk requests. Their intention is to get deep and detailed information.

Appropriate letters for follow up actions will be automatically generated, and the tool will also include questionnaires for users, such as how was the reply, whether it went according to expectations, was the language clear, etc. These questions will be flexible and evolve. Researchers will also experiment with the templates of the requests letter to see how to improve the replies.

The company directory will evolve in stages. They will start populating it with around 100 companies, mainly in the Netherlands but not exclusively. This is carried out by them at present but they would like to open the process, initially with a system for users to correct any company details. They are not planning to provide a full policy analysis, just the basic contact details for a request.






Terms of Service: Didn't Read

<https://tosdr.org>






Terms of Service: Didn't Read (ToS;DR) rates and websites' terms and conditions and privacy policies according to a grading system of "very good – Class A" to "very bad – Class E". The site looks at copyright and ownership issues as well as data protection issues. It refers to these in general as "user rights".

The main functionality of the site is to present grades of websites that assess "the fairness of their terms for users".²⁵ As an example, Class A is "the best terms of services: they treat you fairly, respect your rights and will not abuse your data." Grade E, by contrast, means that "The terms of service raise very serious concerns."






Google Class C

-  Google keeps your searches and other identifiable user information for an undefined period of time
 -  This service tracks you on other websites
 -  Google can use your content for all their existing and future services
 -  Google can share your personal information with other parties
 -  Partial archives of their terms are available
- [More details](#)






YouTube Class D

-  Terms may be changed any time at their discretion, without notice to the user
 -  Processes a personal information (email, id but also device info, location)
 -  The copyright license is broader than necessary
 -  Reduction of legal period for cause of action
 -  They can remove your content at any time and without prior notice
- [More details](#)

GitHub Class B

-  You don't grant any copyright license to github
 -  Changes can happen any time, sometimes without notice
 -  You shall defend and indemnify GitHub
 -  Your personal information is used for limited purposes
 -  Your account can be suspended and your data deleted any time for any reason
- [More details](#)

SoundCloud Class B

-  You stay in control of your copyright
 -  Collected personal data used for limited purposes
 -  6 weeks to review changes
 -  Indemnification from claims related to your content or your account
 -  Personal information can be disclosed in case of business transfer or insolvency
- [More details](#)

²⁵ <https://tosdr.org/classification.html>

The classifications can be contributed by anyone using a bespoke tool made by the project for the purpose of collecting and debating small pieces of analysis about a website.²⁶

Contributors – who have to register and be logged in to the edit service – can add a new website to the database, propose some analysis for a small point of the terms and/or privacy policy, and debate the contributions of others. A curator then has to approve the contributions before it can be made live on the website.

Users can also download browser extensions²⁷ from ToS;DR which provide information about the terms and privacy policies of websites they visit.

The title of the site is a play on words with a popular acronym on the Internet, TL;DR which is short for "Too Long; Didn't Read". People use TL;DR when they write a long piece of text which they expect will not be read by lots of people so they then write a summary after the acronym.

The data files for the analysis of the companies are stored in JSON on Github²⁸ under the Creative Commons Attribution-ShareAlike licence.

As way of a legal disclaimer, the legal information page says that "Nothing here should be considered legal advice. We express our opinion with no guarantee and we do not endorse any service in any way."

PriboT / Polisis

<https://pribot.org>

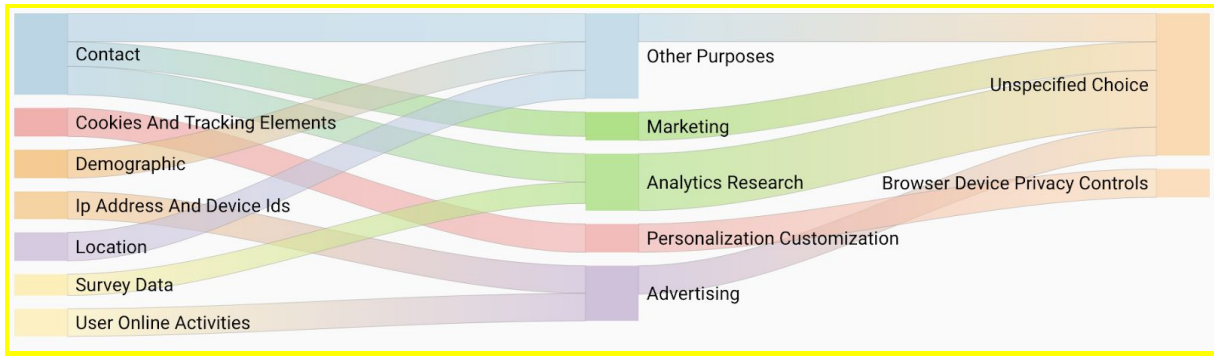
PriboT is a project developed by a group of Swiss and US academics using sophisticated artificial intelligence – deep learning – for the automated analysis of privacy policies. The team analysed over 130,000 privacy policies to build their models.

They have built two tools on the basis of their research. The first one is Polisis, a tool that automatically breaks down the text of privacy policies in categories like data collection – which includes purposes, retention, third party sharing, security or "rights to edit". This information is also translated into privacy icons. In some places, excerpts from the policies are copied verbatim. The platform provides some excellent visual tools to make it easier to understand the results. There is a particularly useful graph that connects data collected with purposes. They have also developed browser plugins to facilitate access to the analysis from the original policy page.

²⁶ <https://edit.tosdr.org>

²⁷ <https://tosdr.org/downloads.html>

²⁸ <https://github.com/tosdr/tosdr.org/tree/master/services>



Their classification follows the taxonomy of the online privacy policies dataset from UsablePrivacy²⁹, where 115 privacy policies were annotated by law students. This classification does not fully match the information required under GDPR, but comes close in some areas, particularly the data collected and purposes. Unfortunately, the rights section is not detailed enough for EU standards.

The researchers claim that Polisis can achieve an accuracy of 88.4% on this automated text classification task, when evaluated against earlier annotations by a group of three legal experts³⁰. This indicates that there is certainly a role for AI in this area, but complementing human intervention.

The other tool based on the same research is a smart agent dialogue bot, Pribot, that allows for unstructured questions to be asked about the policies. The results are somewhat mixed in some unstructured queries and their default questions allow access to the same information available in Polisis in a direct manner. The dialogue interface itself can be useful for particular types of more specific queries, particularly if additional information is required from the user, but less so when the purpose is to navigate a static information tree. A simpler questionnaire could do it more quickly and more accurately presenting multiple choices.

²⁹ Technical report <http://aclweb.org/anthology/P/P16/P16-1126.pdf> and user interface <https://explore.usableprivacy.org/>

³⁰ https://pribot.org/files/Polisis_Technical_Report.pdf