Approaching Access

A comparative analysis of company responses to data access requests in Canada

Citizen Lab Research Brief No. 106

February 12, 2018

[This page intentionally left blank]

Electronic version first published at `https://citizenlab.ca` in 2018 by the authors.

The Citizen Lab is a research group based at the University of Toronto's Munk School of Global Affairs specializing in the intersection between digital technologies and human rights.

Document Version: 1.1

ABOUT THIS REPORT

Document Version: 1.1

Data Access Requests (DARs) enable Canadians to learn more about the kinds of data that organizations collect about them, and what organizations do with this information. *Approaching Access* analyzes datasets derived from DARs between research participants and their telecommunications, fitness tracking, and online dating companies. We compare how companies within particular industries responded to DARs and how different industries compare to one another. We build on this comparison to discuss the strengths and limitations of DARs, and provide recommendations for overcoming their limitations.

This research is led by the Citizen Lab at the Munk School of Global Affairs, University of Toronto. The project was funded via Open Effect by CIRA's 2015-16 Community Investment Program. Additional funding was provided by the Office of the Privacy Commissioner of Canada through its Contributions Program.

SUGGESTED CITATION

Christopher Parsons, Andrew Hilts, and Masashi Crete-Nishihata. 2017. Approaching Access: A comparative analysis of company responses to data access requests in Canada. *Citizen Lab Research Brief No. 106*. Available at: `https://citizenlab.ca/wp-content/uploads/2018/02/approaching_access.pdf`

ABOUT THE CITIZEN LAB

The Citizen Lab is an interdisciplinary laboratory based at the Munk School of Global Affairs, University of Toronto, focusing on research, development, and high-level strategic policy and legal engagement at the intersection of information and communication technologies, human rights, and global security.

`https://citizenlab.ca`

ABOUT THE AUTHORS

Christopher Parsons received his Bachelor's and Master's degrees from the University of Guelph, and his Ph.D from the University of Victoria. He is currently a Research Associate at the Citizen Lab at the Munk School of Global Affairs as well as the Managing Director of the Telecom Transparency Project at the Citizen Lab.

Andrew Hilts is a Senior Researcher and Developer at the Citizen Lab at the Munk School of Global Affairs. His research and software development focuses on empowering citizens to exercise their digital rights online.

Masashi Crete-Nishihata is Research Director at the Citizen Lab, Munk School of Global Affairs, University of Toronto. He researchers the socio-political impact of information controls.

ACKNOWLEDGEMENTS

Thank you to Adam Senft and Bram Abramson for review and copyediting. We are grateful to Ron Deibert for research guidance and supervision. This research would not have been possible without the Access My Info users who participated in this study.

CONTENTS

EXECUTIVE SUMMARY

When someone makes a phone call they may not be thinking about how the location of their mobile phone is recorded based on the address of the cell tower to which the phone connects. A person looking for love may provide photos, chat messages, and a history of interpersonal connections to an online dating service, without thinking about how that data might be stored even after they delete their account. And a user of a fitness tracker might be happy to send detailed workout statistics to a fitness app company so that they can compete against their friends on fitness goals, but be unaware about how that data is being protected against unauthorized access.

Without knowing who is collecting personal data, for what purpose, or for how long, or the grounds under which they share it, a consumer cannot exercise their rights nor evaluate whether an organization is appropriately handling their data. Canada's commercial privacy legislation, the Protection of Personal Information and Electronic Documents Act (PIPEDA), empowers Canadians to issue legally-binding Data Access Requests (DARs) to private companies to answer exactly these kinds of questions. This report is the result of a three year study of DARs in Canada that shows what happens when telecommunications companies, fitness trackers, and online dating services are asked by consumers to provide transparency into their data privacy practices and policies.

Between 2014-2016 we recruited participants to systematically issue DARs to telecommunications companies, fitness trackers, and online dating services used by Canadians to evaluate a series of research questions:

- What proportion of companies contacted would respond to DARs at all?
- What proportion of companies that did respond to DARs would respond in a relatively complete manner to all questions asked?
- What proportion of companies that did respond to DARs would provide individuals with copies of their personal information at no or minimal cost?
- What commonalities or differences would be found in responses to individuals in each industry group studied, and across industries?
- To what extent would individuals who received responses be satisfied with the information they received and what, if anything, might be done to improve organizations' disclosures to enhance individuals' satisfaction?

INCONSISTENT RESPONSES ACROSS COMPANIES AND INDUSTRIES

Participants received responses from companies but the information provided varied widely across companies and industries. Variations included:

- the specificity with which requester questions are answered;
- what types of data are returned;
- whether or not data retention periods are published; and
- clarity about data disclosures to third parties, including government authorities.

BARRIERS TO ACCESS

Participants also encountered barriers to accessing the private information that companies retained about them. These barriers included:

- identity verification procedures;
- secure data transfer requirements;
- costs offloaded to requesters; and
- push-back by some non-Canadian companies as to whether their services to Canadian consumers in Canada are, in fact, bound by Canadian privacy law.

TOWARDS IMPROVED DATA ACCESS IN CANADA

Our report concludes with recommendations for how businesses can improve their DAR processes and related data transparency efforts, and allow citizens to more effectively exercise stewardship over their personal data. We make seven key recommendations:.

- Companies should prepare and produce data retention schedules that identify specific types of information they collect and the period of time for which they retain it.
- Companies should prepare and publish government access handbooks that identify the different kinds of personal information they hold, and establish the specific legal powers and processes to be undertaken before the company will disclose a subscriber's personal information.
- Companies should prepare transparency reports that disclose the regularity, and rationale for which, government agencies request access to subscriber-related information.
- Companies should collaborate within their respective industries to establish common definitions for personal data mini-collections to which common policies are applied, such as subscriber data, metadata, content of communications, etc.

- Companies should not assume they know which communications method their customers would prefer to use when discussing a DAR letter. They should first ask the customer what their preferred method is, and only then pose questions to clarify the requester's inquiries.
- Companies should publish data inventories describing all the kinds of personal information that they collect, and freely provide copies of a small set of representative examples of records for each kind of personal information to subscribers upon request.
- Either individual organizations or industry groups should communicate with non-corporate stakeholders to help streamline the request process, or to help establish requesters' expectations. This effort might involve developing Application Programming Interfaces (APIs) to expedite the issuance and response to DAR letters, or working to modify language used by web applications to more accurately reflect the data that might be handled by organizations in the course of commercial activity.

DARs provide a valuable method for understanding the kinds of information which are collected, retained, processed, and handled by private companies. This report provides a look at how companies respond when Canadians exercise their access rights. It also draws lessons from within specific industry groupings as well as across industries. Given the amounts of digital information that individuals confide to third parties on a daily basis it is imperative that they can gain access to such information upon request, especially when companies do not publish clear guidance as to their broader data collection, retention, handling, or disclosure practices. Our report showcases how DARs can provide insight into corporate practices. But, at present processes surrounding DAR-handling and -processing are immature. Advancing DAR practices and policies requires either private-sector coordination to advance individuals' access to their personal information, or regulatory coordination to clarify how private organizations ought to provide access to the information of which they are stewards.

1 INTRODUCTION

For the past three years Open Effect and the Citizen Lab have collaborated to help people better understand how companies in a variety of industries collect personal information, what is done with that collected information, with whom it is shared, and – essentially – to enable people to make requests for their own personal information. This process began as an effort to understand some of the activities undertaken by Canadian telecommunications companies but, since then, has grown to include online dating companies that provide services in Canada, and many of the largest fitness tracker-producing companies in the world.[1]

This study has revealed gaps between privacy legislation in theory and in practice. Despite access to personal information being a bedrock of personal information protection legislation, and having existed as law for over a decade in Canada, we found that companies routinely lack mature policies to help customers understand what information is collected, processed, and used in the course of providing services to them.

This report analyzes company responses to Canadian customers' requests for access to their personal information ("Data Access Requests" or DARs).[2] This research builds on past projects on telecommunications companies' practices in Canada,[3] and privacy practices and policies of fitness trackers.[4]

We investigated DAR responses sent by research participants to leading companies in the telecommunications, fitness tracking, and online dating industries. We undertook industry-specific as well as cross-industry analyses to ascertain whether data provided in company responses substantially addressed the questions raised in the DARs, as well as instances where different protocols are needed to more effectively and completely respond to individuals' requests.

---

1 This project has also been adopted internationally, including in Hong Kong to enable local residents to make data access requests to their local telecommunications providers.

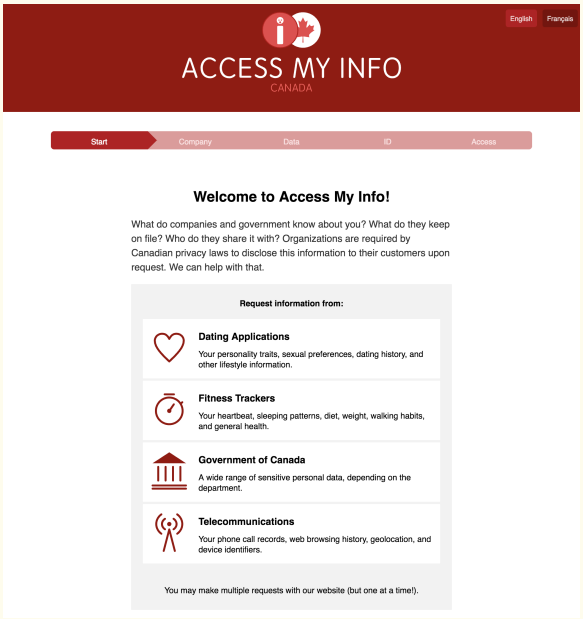
2 Data Access Requests also have other terms, in other jurisdictions, including Subject Access Requests (SARs) in Europe.

3 Parsons, Christopher. (2015). "The Governance of Telecommunications Surveillance: How Opaque and Unaccountable Practices and Policies Threaten Canadians," *The Citizen Lab*, retrieved February 15, 2017, `http://www.telecomtransparency.org/wp-content/uploads/2015/05/Governance-of-Telecommunications-Surveillance-Final.pdf`.

4 Hilts, Andrew; Parsons, Christopher; and Knockel, Jeffrey. (2016) "Every Step You Fake: A Comparative Analysis of Fitness Tracker Privacy and Security," *The Citizen Lab and Open Effect*, retrieved February 15, 2017, `https://openeffect.ca/reports/Every_Step_You_Fake.pdf`.

Access My Info

Open Effect and the Citizen lab developed Access My Info (AMI), a web application[a] that makes it easy for Canadians to create requests for access to their personal data. AMI presents website visitors with a step-by-step wizard that helps users to create a personalized data access request.



The Access My Info landing page.

The letter generated using the tool can be saved as a PDF for emailing, printed and mailed to a company, or in some cases emailed directly to a company's privacy officer. The AMI web application does not send requests to an organization on users' behalf: individuals must send their own requests. Though this slightly increases the time commitment of users it also ensures that requests are clearly sent by individuals, as opposed to a web tool's provider. The intent, in this methodology, was to reduce the likelihood that companies dismiss requests as vexatious.[b]

[a] Access My Info is publicly available at `https://accessmyinfo.org`

[b] Hilts, Andrew (2016). "Access My Info Software Design Document," Open Effect, retrieved December 7, 2017, `https://openeffect.ca/wp-content/uploads/2017/01/ami-design-doc.pdf`.

We found significant variation in how companies responded to DARs both within and across the studied industries. We identified several barriers in the DAR process that can serve to discourage requesters and stymie access to their personal information. We present recommendations for how companies can improve their DAR processes and arguably improve the trust their

customers place in them.

We ultimately argue, however, that responses to DARs are limited in what they can reveal about a company's data handling practices. A greater degree of insight into company data practices can be achieved through research that compares DAR responses to technical analysis of data flows, corporate privacy policies, and external documents, such as those held by law enforcement agencies.

1.1 DIVISION OF REPORT

2. Background and Research Questions
This section provides a background to access to personal information law in Canada and a more comprehensive explanation of the report's research questions.

3. Data Collection and Methodology
This section describes our data collection activities and research methodology.

4. Comparative Analysis of Company Responses
This section presents our comparative analyses of company responses. We examine the telecommunications, online dating, and fitness tracking industries.

5. Comparison of Responses Across Industries
This section presents a cross-industry analysis of our three studied industries.

6. Participant Reflections on Company Responses
This section presents the results of a survey we conducted with people who made DARs.

7. Limitations of Data Access Request Practices
This section describes some of the limitations of relying on DARs to fully understand company practices around personal information.

8. Recommendations to Improve Data Access Requests
This section offers recommendations to companies to improve the transparency of their data collection and best practices for responding to DARs.

9. Future Work
This section outlines possible avenues of future work linked to DARs.

10. Conclusion

This section presents a summary of the key points raised in the report.

2 BACKGROUND AND RESEARCH QUESTIONS

This section provides a background to access to personal information law in Canada and a more comprehensive explanation of the report's research questions.

2.1 ACCESS TO PERSONAL INFORMATION

Data protection legislation routinely includes rights of access to personal information.[5] Such rights, in their most basic form, are designed to let consumers file data access requests (DARs) to better understand what information private companies retain about them in order to correct erroneous data.[6] Without knowing who is collecting personal data, for what purpose, or for how long, or the grounds under which they share it, a consumer cannot exercise their rights nor evaluate whether an organization is appropriately handling their data.[7] In the European Union, data access rights have been linked to human rights: everyone has a right of access to information that is collected about them.[8] In other jurisdictions such access rights are principally linked with consumer rights, and only antecedently linked with broader constitutional principles or rights.

Access rights vary by jurisdiction. In some jurisdictions, individuals can send DARs for all of the personal information about them held or processed by the corporation that holds the data. In other jurisdictions, individuals and corporations will negotiate the scope of the data provided to satisfy the requester's concerns. And in other cases, data may be presented visually or described verbally without providing a material copy. Compounding the differences across jurisdictions are the costs of DARs. Where laws are premised on consumer protection laws that authorize the correction of incorrect information, costs are usually intended to be low on the basis that relatively few records will be requested. But costs for a comprehensive DAR may involve material charges to the requester, depending on the intent of the data access legislation and enforcement of this intent as against charges imposed.

---

5 Bygrave, Lee Andrew . (2014). *Data Privacy Law*. Oxford: Oxford University Press; Bennett, Colin J. and Raab., Charles D. (2006). *The Governance of Privacy: Policy Instruments in Global Perspective*. Cambridge, Mass.: The MIT Press.

6 Hunter, L. (2003). "Hand over the Information-It's Mine, It's Personal," *LawNow* 28 (April/May); European Commission (Justice) (2011). "Protecting your data: your rights," European Commissioner, retrieved December 6, 2017, `http://ec.europa.eu/justice/data-protection/individuals/rights/index_en.htm`; Office of the Privacy Commissioner of Canada. (2017). "Accessing your personal information," *Office of the Privacy Commissioner of Canada*, retrieved December 6, 2017, `https://www.priv.gc.ca/en/privacy-topics/access-to-personal-information/accessing-your-personal-information/`.

7 Gellman, Robert. (2017). "Fair Information Practices: A Basic History," *SSRN*, retrieved December 6, 2017, `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2415020`.

8 European Union, Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended), Article 8.

In Canada, consumer data access rights are enshrined in the Personal Information and Protection of Electronic Documents Act (PIPEDA) which, in 2000, entrenched a set of principles set down in 1995 by the Canadian Standards Association (CSA)'s *Model Code for the Protection of Personal Information*[9], itself based on the Organization for Economic Cooperation and Development (OECD)'s 1980*Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*.[10] The CSA's *Model Code for the Protection of Personal Information* was ultimately approved by the Standards Council of Canada, the Crown corporation which promotes standards development, promotion and implementation, as a "National Standard of Canada". It was constructed around ten principles and accompanying commentary and detailed in a 1997 implementation workbook.[11] Of the ten CSA and, now, PIPEDA principles, Principle 4 ("Limiting Collection") relates most directly to DARs, as its Clause 4.9 requires that, "[u]pon request, an individual shall be informed of the existence, use, and disclosure of his or her personal information and be given access to that information"[12] provided the subject matter, parties, or territory in question have a real and substantial connection to Canada such that PIPEDA has jurisdiction.[13] In addition, "an individual shall be able to challenge the accuracy and completeness of the information and have it amended as appropriate."[14]

The CSA *Model Code*'s original fourth principle focused on enhancing consumers' understanding of data collection, handling, and disclosure policies and enabling those affected to remedy incorrect or superfluous information. The access right in PIPEDA, correspondingly, was meant to let individuals correct records about themselves, such as credit history information, purchase logs, or other information that might have an inappropriate impact on consumers' lives.

It can be challenging for individuals to know how an organization collects and handles personal information based only on publicly-available corporate documentation, such as a privacy policy,[15] and so the PIPEDA-backed access right can also be used for insight into a company's

---

9 Bennett, Colin J. and Raab, Charles D. (2006). *The Governance of Privacy: Policy Instruments in Global Perspective*. Cambridge, Mass.: The MIT Press; Canadian Standards Association (CSA). (1996). *Model Code for the Protection of Personal Information*, CAN/CSA-Q830-96, CSA, Rexdale, in Bennett and Raab 2006.

10 Organization for Economic Cooperation and Development. (1980). *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, retrieved November 16, 2017, `https://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm`.

11 Canadian Standards Association (CSA). (1997). *Making the Privacy Code Work For You*, CSA, Rexdale, in Bennett and Raab 2006.

12 Office of the Privacy Commissioner of Canada. (2013). "Interpretation Bulletin: Access to Personal Information," *Office of the Privacy Commissioner of Canada*, last modified May 16, 2013, retrieved September 23, 2014, `https://www.priv.gc.ca/leg_c/interpretations_05_access_e.asp`.

13 *Lawson v. Accusearch Inc*., 2007 FC 125; *A.T. v. Globe24h.com*, 2017 FC 114, paragraph 53.

14 Personal Information Protection and Electronic Documents Act (PIPEDA), S.C. 2000, c. 5, Schedule 1, Clause 4.9..

15 Office of the Privacy Commissioner of Canada. (2013). "Backgrounder: Results of the 2013 Global Privacy Enforcement Network Internet Privacy Sweep," *Office of the Privacy Commissioner of Canada*, August 13, 2013, retrieved September 23, 2014, `https://www.priv.gc.ca/media/nr-c/2013/bg_130813_e.asp`; McDonald, A. M., & Cranor, L. F. (2008). "The cost of reading privacy policies," *I/S: A Journal of Law and Policy for the Information Society 4*; McDonald A.M., Reeder R.W., Kelley P.G., Cranor L.F. (2009) "A Comparative Study

data-handling practices.[16] As a result, consumers can use a right focused on fixing incorrect recordkeeping to expand their understanding of an organization's ongoing collection of personal information over time. The right's exercise may also clarify vague or confusing elements of privacy policies or terms of service pertaining to how personal information is collected, processed, and disclosed to other parties.

Research in other jurisdictions has revealed data access challenges. A study of DARS issued in Hong Kong found variation among telecommunications companies in the DAR response process, the definition of personal data, and whether or not IP addresses were considered personal data.[17] Recent work looking at DARs in the European Union have found requesters face many different barriers to obtaining fulsome responses to their DARs.[18] Other EU-based research found significant variation around how satisfied requesters were with responses to their access requests, and generally found them to be inadequate in providing deep insight into companies' data practices.[19] Our report explores what sort of barriers and variation exist for DARs sent within the Canadian context.

2.2 INDUSTRIES ANALYZED

We analyze three industries in this report: telecommunications, fitness tracking, and online dating.

Telecommunications companies were included as part of a broader Citizen Lab project to determine the data-handling practices of wireline and wireless telecommunications providers. A data access methodology was adopted because privacy policies and company statements did not provide detailed information about corporate practices around retention and use of per-

of Online Privacy Policies and Formats." In: Goldberg I., Atallah M.J. (eds). Privacy Enhancing Technologies. *PETS 2009*. Lecture Notes in Computer Science, vol 5672. Springer, Berlin, Heidelberg; Obar, Jonathan A. and Oeldorf-Hirsch, Anne. (2016).," The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services," *TPRC 44: The 44th Research Conference on Communication, Information and Internet Policy 2016*.

16 Hilts, Andrew and Parsons, Christopher. (2014). "Enabling Citizens' Right to Information in the 21st Century," *The Winston Report*, Fall 2014; Hilts, Andrew; Parsons, Christopher ; and Knockel, Jeffrey. (2016). "Every Step You Fake: A Comparative Analysis of Fitness Tracker Privacy and Security," *The Citizen Lab and Open Effect*, retrieved December 4, 2017, `https://openeffect.ca/reports/Every_Step_You_Fake.pdf`; Bennett, Colin J.; Parsons, Christopher; and Molnar, Adam. (2014). "Real and Substantial Connections: Enforcing Canadian Privacy Laws Against American Social Networking Companies," *Journal of Law, Information & Science 23(1)*.

17 Hargreaves, Stuart and Tsui, Lokman, (2017). IP Addresses as Personal Data Under Hong Kong's Privacy Law An Introduction to the Access My Info HK Project. *Journal of Law Information, and Science 25*.

18 Norris, C., de Hert, P., L'Hoiry, X., and Galetta, A. *The Unaccountable State of Surveillance: Exercising Access Rights in Europe*; Thompson, Barney. (2018). "Amazon and Facebook fare badly in personal data test". *Financial Times*, `https://www.ft.com/content/5c1987d2-05d2-11e8-9650-9c0ad2d7c5b5`.

19 Mahieu, Rene and Asghari, Hadi and van Eeten, Michel. (2017) "Collectively Exercising the Right of Access: Individual Effort, Societal Effect." *GigaNet (Global Internet Governance Academic Network) Annual Symposium 2017*. Available at SSRN: `https://ssrn.com/abstract=3107292`.

sonal information.[20] We saw DARs as a possible solution to addressing this knowledge gap.[21] The Access My Info (AMI) web application[22] was initially developed to support research into these companies' practices.

Fitness tracker companies were included as an example of how consumer 'Internet of Things' (IoT) devices collect and secure information. Fitness trackers collect data about their wearer's heart rate, steps, calories burned, sleep patterns, height, weight, fitness goals, diet, and more. This data is used to present users with a window into their personal fitness. In some cases, users can share some of their fitness data over the Internet with their friends, to compete, hold one another accountable, and congratulate one another on achieving fitness milestones.

In previous work we analyzed security and privacy issues with fitness trackers through technical investigation, policy analyses and filing DARs in an attempt to learn what data the respective companies had collected about their users.[23] In this report we focus on what was learned through the filing of DARs.

Online dating companies often collect extensive amounts of highly-detailed personal information. This data includes information related to individuals' finances, personal health, religion, lifestyle, hobbies, sexual preferences, mental health, chat, and geographical information, among others. Online dating services are reportedly used by more than a third of Canadians as of 2015.[24] Users upload intimate photos, messages, and profile details, which are stored on dating app servers. The privacy interest in dating applications is clear: A 2014 Pew Research Centre survey found that 71% of Americans regard their relationship history as very or somewhat sensitive data.[25]

There have been journalistic accounts of requesting data from online dating companies in European jurisdictions,[26] but no systematic analysis of how companies in this industry respond to DARs.

---

20 Parsons, Christopher (2014). "Responding to the Crisis in Canadian Telecommunications," *The Citizen Lab*, retrieved December 3, 2017, `https://citizenlab.org/2014/05/responding-crisis-canadian-telecommunications/`.

21 Parsons, Christopher. (2015). "Beyond the ATIP: New methods for interrogating state surveillance." In Jamie Brownlee and Kevin Walby (Eds.). *Access to Information and Social Justice*. Arbeiter Ring Publishing.

22 Access My Info is a web application designed to make it easier for people to create requests for access to their personal information.

23 Hilts, Andrew; Parsons, Christopher; and Knockel, Jeffrey. (2016). "Every Step You Fake: A Comparative Analysis of Fitness Tracker Privacy and Security," *The Citizen Lab and Open Effect*, retrieved February 15, 2017, `https://openeffect.ca/reports/Every_Step_You_Fake.pdf`.

24 Thottam, Isabel. (2017). "10 Online Dating Statistics You Should Know," *eHarmony*, retrieved Decemer 6, 2017, `https://www.eharmony.ca/online-dating-statistics/`.

25 Pew Research Centre (2014). "Americans Consider Certain Kinds of Data to be More Sensitive than Others," *Pew Research*, retrieved December 7, 2017, http://www.pewinternet.org/2014/11/12/americans-consider-certain-kinds-of-data-to-be-more-sensitive-than-others/.

26 E.g. Duportail, Judith. (2017). "I asked Tinder for my data. It sent me 800 pages of my deepest, darkest secrets." *The Guardian*. `https://www.theguardian.com/technology/2017/sep/26/tinder-personal-data-dating-app-messages-hacked-sold`

2.3 POLICY RATIONALES FOR STUDY

Companies that provide telecommunications, fitness tracking, and online dating services are collecting, processing, managing and, potentially, disclosing significant amounts of customer data. Such activities are sometimes undertaken directly by the companies with which Canadians voluntarily entered into a subscriber relationship. At other times they are undertaken by third parties on behalf of those companies. The aggregate result is that the number of companies that consumers know are processing their data is growing, but the number of companies processing their data that are unknown to the consumers is also growing, perhaps even more quickly.[27] Neither privacy policies, nor other corporate literature, necessarily disclose to a customer what data in particular is collected, for how long they are retained, or which companies might subsequently process it.[28] Not only is such information often absent in public-facing corporate documentation, but the privacy policies themselves can be vague and challenging to understand, even for trained professionals.[29] Furthermore, unless customers review companies' privacy policies on a regular basis, they may never be aware that the policy that existed when they first joined a service has subsequently changed.

Even if a privacy policy or other public facing corporate document delineates the information collected, how long it is stored for, why it is processed and who is responsible for that processing, it is possible that the document is incorrect or that the organization has collected inaccurate information. Data access rights address such a situation and give consumers an opportunity to verify that the documents they read accurately represent a company's actual processes. Past research has shown that these public-facing documents do not always capture all data-handling or processing activities. Companies may change practices and fail to update such documents. The documents themselves may make claims about data processing that do not align with ac-

27 Solove, Daniel. (2004). *The Digital Person: Technology and Privacy in the Information Age*. New York: New York University Press; Hoofnagle, C. J. (2003). "Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement," *NCJ Int'l L & Com Reg*.

28 Vu KP.L. et al. (2007). "How Users Read and Comprehend Privacy Policies." In Smith M.J., Salvendy G. (eds). *Human Interface and the Management of Information: Interacting in Information Environments. Human Interfac*e. Lecture Notes in Computer Science, vol 4558. Springer, Berlin, Heidelberg; Jensen, C., & Potts, C. (2004). "Privacy policies as decision-making tools: an evaluation of online privacy notices," New York, New York, USA: ACM. http://doi.org/10.1145/985692.985752; Protalinski, Emil. (2012). "Survey: Facebook, Google privacy policies are incomprehensible," *ZDNet*, retrieved December 6, 2017 `http://www.zdnet.com/article/survey-facebook-google-privacy-policies-are-incomprehensible/`; Dwoskin, Elizabeth. (2015). "Privacy Policies More Readable, But Still Hard to Understand," *Wall Street Journal*, retrieved December 7, 2017, `https://blogs.wsj.com/digits/2015/12/30/privacy-policies-more-readable-but-still-hard-to-understand/`.

29 Hochhauser, M. (2001). "Lost in fine print: Readability of financial privacy notices," *Privacy Rights Clearinghouse*, retrieved December 7, 2017, `https://www.privacyrights.org/blog/lost-fine-print-readability-financial-privacy-notices-hochhauser?page=7`; Peslak, A. R. (2005). "Privacy policies of the largest privately held companies: a review and analysis of the Forbes private 50," *Proceedings of the 2005 ACM SIGMIS CPR conference on Computer personnel research*, Atlanta, Georgia, USA, Session 5.2: Organizational policies and practices, 104 – 111;

tual practices. Technical mistakes or poor inter-departmental communication internally may cause a company's activities to deviate from those explained to customers.[30]

Ultimately, then, the rationale for this line of research is threefold:

1. To help individuals understand the agreements they consented to and which govern the collection and use of their data and personal information. Sometimes this inquiry entails correlating bland statements of categories of data collected against actual records provided by a company following a DAR, or identifying the full scope of data being collected in the first place.
2. This line of research showcases the relative value of helping consumers try to correct records when the information they receive has been incorrectly captured by the companies in question.
3. This research and its associated tools for facilitating DARs help educate the public that they possess this right and help them more readily exercise the right.

2.4 RESEARCH QUESTIONS

Access My Info (AMI) DARs were meant to help researchers, and the public more broadly, develop a deeper understanding of how personal information is collected, what it is used for, how long it is kept, and with whom it is shared.

Beyond specific data retention questions associated with each discrete industry segment (telecommunications, fitness tracking, dating) researched, there were also secondary questions as to DARs' very validity. Although backstopped by Canadian law and paralleled in other jurisdictions, we hypothesized based on past research that companies would not provide fulsome responses to DARss. For this report, we focus our analyses exclusively on the process, and data resulting from, DAR filings. We asked:

- What proportion of organizations contacted would respond in any way to DARs?
- What proportion of organizations that did respond to DARs would provide relatively complete responses to all questions asked?
- What proportion of organizations that did respond to DARs would provide individuals with copies of their personal information at no or minimal cost?
- What commonalities or differences would be found in responses to individuals in each industry group studied, and across industries?

[30] Bennett, Colin; Parsons, Christopher; Molnar, Adam. (2014). "Forgetting and the right to be forgotten." In Serge Gutwirth et al. (Eds.), Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges, Springer.

- To what extent would individuals who received responses be satisfied with the information they receive and what, if anything, might be done to improve organizations' disclosures to enhance individuals' satisfaction?

Finally, we wanted to understand the efficacy of DARs more broadly as a method for better understanding corporate personal information collection, data minimization, and third-party disclosures.

3 DATA COLLECTION AND METHODOLOGY

This section describes our data collection activities and research methodology.

Our primary data is correspondence relating to data access requests (DARs) filed by our research participants with their telecommunications, fitness trackers, or online dating service providers. The correspondence includes initial request letters and all subsequent written interactions between company and requester as well as any data attachments included in such interactions.

3.1 DATA SETS

The data sets analyzed in this report includes some which were collected in previous research reports on telecommunications companies (2014) and fitness tracking companies (2015-16), as well as original data collected on telecommunications and online dating companies (2016).

3.1.1 TELECOMMUNICATIONS PROVIDERS

Our telecommunications data collection in 2014 and 2016 examined Canadian telephone, mobile, and Internet service providers. We used a mixed-methods approach to collecting data, which involved collating available information on corporate activities and their regulation, as well as submitting DARs. We examined available case law, legislation, scholarly articles, and decisions by the Office of the Privacy Commissioner of Canada to understand the contours of how telecommunications companies collected data and their obligations to inform customers as to how these providers collected, processed, stored, and secured the collected information. The approach also involved discussions with some telecommunications industry and government stakeholders to understand the broader context of government requests for data held or collected by telecommunications companies, actively asking companies to explain their data collection and handling practices, and filing access to information requests with government organizations to understand government's role in requesting information from telecommunications providers.[31]

3.1.2 FITNESS TRACKERS

Our fitness tracker research (2015-16) involved analysing privacy policies and technical research and reviewing personal data request responses by the companies included in the project. There

[31] For a detailed accounting of methods used for this case, see: Parsons, Christopher. (2015). "Beyond the ATIP: New Methods For Researching State Surveillance Techniques." In Jamie Brownlee and Kevin Walby (Eds.).*Access To Information And Social Justice: Critical Research Strategies for Journalists, Scholars, and Activists*. Winnipeg, Manitoba: ARP Books.

was also limited engagement with select companies when security vulnerabilities were found as a result of technical research.[32] Desk research included examining case law, legislation, scholarly articles, and decisions by privacy commissioners.

3.1.3 ONLINE DATING

The third project (2016-2017) relied exclusively on DARs to gauge how and why online dating companies collected and shared personal information and for how long they retained it. Desk research included examining case law, legislation, scholarly articles, and decisions by privacy commissioners.

3.2 DEVELOPMENT OF THE DATA ACCESS REQUEST QUESTIONS

Each company we studied was sent a letter based on a common baseline set of questions adapted from an earlier project by Colin Bennett and Christopher Parsons, 'Canadian Access To Social Media Information (CATSMI)', which identified 33 kinds of questions that could be used to analyze privacy policies and five broad questions when issuing personal data requests to private companies.[33]

CATSMI was focused on what kinds of information social media companies collected about their users, the ease with which customers could obtain this information, and the kinds of information government agencies could compel from these companies by comparison.[34] Here, a common core of questions was adjusted slightly to accommodate each of the three industry segments. Questions to telecommunications providers focused on the retention of data and metadata generated in the course of using telecommunications services. Slightly different data was sought from fitness tracking and online dating services. Privacy law experts were consulted to ensure questions complied with PIPEDA and the broader spirit of Canadian law, again to reduce the likelihood that questions would be rebuffed or challenged.

3.3 PARTICIPANT RECRUITMENT

Research participants were recruited in two pools: pilot volunteers and AMI tool users.

---

[32] Hilts, Andrew; Parsons, Christopher; and Knockel, Jeffrey. (2016). "Every Step You Fake: A Comparative Analysis of Fitness Tracker Privacy and Security," *The Citizen Lab and Open Effect*, retrieved February 15, 2017, `https://openeffect.ca/reports/Every_Step_You_Fake.pdf`.

[33] See: `https://christopher-parsons.com/catsmi-project/`.

[34] Bennett, Colin; Parsons, Christopher; Molnar, Adam. (2014). "Forgetting and the right to be forgotten." In Serge Gutwirth et al. (Eds.). *Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges*. Springer. Bennett, Colin; Parsons, Christopher; Molnar, Adam. (2014). "Real and Substantial Connections: Enforcing Canadian Privacy Laws Against American Social Networking Companies," *Journal of Law, Information & Science*.

Year	Industry	Participants
2014	Telecommunications	6
2015-16	Fitness Trackers	8
2016	Telecommunications	5
2016	Online Dating	5

Table 1: Overview of research participants across industry data sets

Pilot Volunteers: For each data set, we recruited a group of participants to make DARs to specific companies. We sought to recruit participants for each major company in each industry segment, but were unable to find participants to make requests of some companies.

AMI Tool Users: A broader group of participants was recruited through the AMI web application. After users created their DAR using the tool, they could opt into being contacted about their AMI experience. These users were then asked to share information about the DAR responses they had received, if any; their opinions about the tool's usability; and whether they believed companies had fulsomely responded to the questions provided.

Table 1 provides a breakdown of the participants in each industry sector across data sets.

Appendix A includes copies of the questions that individuals could send to companies using either a manual process or generated by AMI. Appendix B includes copies of the survey questions sent to the broader pool of users who chose to be contacted by researchers.

4 COMPARATIVE ANALYSIS OF COMPANY RESPONSES

In this section we provide an industry-level analysis of the DAR responses telecommunications, fitness tracking, and online dating companies provided to our research participants. We identify commonalities and significant differences in how companies in different industry sectors responded to Canadian citizens' and residents' data access requests. We have published the full data set of our results on Github.[35]

4.1 TELECOMMUNICATIONS PROVIDERS

In both our 2014 and 2016 data sets, all telecommunications companies that were issued DARs provided responses. The responses we analyzed answered the questions in the DARs in varying ways. Data provided included customer service logs, and most telecommunications providers stated they would require payment in exchange for access to detailed technical metadata.

4.1.1 WHAT INFORMATION IS COLLECTED

Service providers do not have uniform data collection policies across the industry. Different services attract different sorts of data collection than others, and practices between companies offering comparable services can also vary. One motivation for customers' DARs is that they cannot understand these practices based on public documents provided by the company in question. DARs can be used to compel companies to clarify what types of information are collected about their customers. Table 2 presents an overview of the information provided by telecommunications service providers in 2014 and, in Table 3, in 2016. A list of the questions sent to each provider is set out in Appendix A.

SUBSCRIBER DATA

Canadian telecommunications providers have historically engaged in controversial warrantless disclosures of 'subscriber data' to requesting government agencies. Canadian telecommunications companies' transparency reports indicate that such warrantless disclosures have plummeted and, in many cases, no longer occur following a ruling by the Supreme Court of Canada.[36] However, policy debates have continued since that ruling to determine whether there are conditions under which subscriber data might be released to government agencies. The past and the current debates have routinely seen shifting definitions of what, specifically, subscriber data itself consists of.

[35] Approaching Access: DAR Analysis. *Citizen Lab*. `https://github.com/citizenlab/approaching-access-data/blob/master/dar-analysis-data.pdf`.

[36] R. v. Spencer, [2014] 2 S.C.R. 212.

Company	Request Date	First response	Notes on metadata access
Fido	2014-07-09	2014-08-08	IP logs / SMS metadata available for a fee, asks requester to provide time period to get cost estimate. Requester did not follow up.
Koodo	Spring 2014	Spring 2014	After pushback from requester, company says full IP address records would take 60 hours at $20/hour (totalling $1,200). Requester did not follow up.
Northwest-Tel	Unknown	2014-08-12	No indication of retention periods for IP logs or SMS metadata. Did not offer to provide cost estimate or mention that customer could get access. Requester did not follow up.
Primus	2014-05-08	2014-05-26	One year of DSL history with IP addresses provided at no cost.
Rogers	2014-05-05	2014-06-05	IP logs and SMS metadata available for a fee, asks requester to provide time period to get cost estimate. Requester did not follow up.
TekSavvy	2014-05-02	2014-06-02	Stated IP address retention period is 30 days after termination of IP lease.

Table 2: 2014 overview of telecommunications provider DARs

Company	Request Date	First response	Notes on metadata access
Bell	2016-09-30	2016-11-29	Mentioned it will take significant resources and time to create logs of previous IP addresses, asked requester to specify a time period in order to get a cost estimate. States that call routing information is not personal information. Requester not believed to have followed up.
Fido	2016-08-26	2016-09-09	Mentioned SMS and cell site data could be obtained for a fee. After significant back and forth, requester paid $100 plus tax for one month's SMS metadata, and $100 plus tax for one month's cell site data.
Rogers	2016-08-26	2016-09-08	Mentioned company does not collect IP addresses or domain names of sites visited. GPS information only collected when sending or receiving a phone call or text message. Call log information available at a price of $15/month, for logs older than 18 months. SMS metadata available only for a fee.
Shaw	2016-06-23	2016-08-18	Archived IP address history can be retrieved at the cost of $250 per year and per modem. Participant did not follow up.
WIND	2016-06-22	2016-07-18	SMS metadata: mentioned that it may be retained for network operations and compliance with law enforcement requests, but did not mention that customer could get access. Participant did not follow up.

Table 3: 2016 overview of telecommunications provider DARs

The 'subscriber data' that is currently disclosed to a government agency, warrantlessly or otherwise, varies by court order and by what telcos themselves retain in their customer records systems. Sometimes a subscriber data demand involves a telecommunications company only releasing a customer's name and address (so-called "CNA," for "customer name and address," requests). In other cases a 'subscriber data' request can be broader and include items such as email addresses, financial information, or other pieces of information.

DAR responses revealed significant variations in what telecommunications providers themselves considered 'subscriber data': some included government identification information, dates of birth, social insurance numbers, email addresses, or credit card information. That the industry did not have a common definition for what this category includes suggests that public confusion as to whether subscriber data refers to CNA information, or more extensive collection of information, is warranted.

METADATA

Some of the most sensitive information disclosed to telecommunications service users relates to their metadata, which is information about their communications. This kind of information can be used to trace where individuals have physically been, where they have visited online, with whom they have communicated, and more.[37]

Telecommunications providers did not generally provide detailed records in response to participant requests for call logs, cell tower connections, and other metadata. Some companies informed requesters that they could provide records for a fee if the requester specified a date range. Companies tended to not offer requesters a sample set of records prior to receiving a fee.

The template of DAR questions we submitted addressed geolocation. Responding companies indicated they retained geolocation coordinates after a communication had taken place, but only Koodo indicated that cell tower information was collected as the result of a user's mere connection to a tower. WIND wrote that they would disclose location information in the instance of either an e911 call or government agency request to track a person's phone using GPS. A participant who made requests of Fido received, after communicating with several different Fido

[37] Office of the Privacy Commissioner of Canada. (2014). "Metadata and Privacy," *Office of the Privacy Commissioner of Canada*, retrieved July 6, 2017, https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2014/md_201410/; Strandburg, Katherine. (2008). "Surveillance of Emergent Associations: Freedom of Associations in a Network Society." In A. Acquisti and S. Gritzalis (eds). *Digital Privacy: Theory, Technologies, and Practices*. New York: Auerbach Publications, pp. 435-458; Parsons, Christopher; Israel, Tamir. (2016). "Gone Opaque? An Analysis of Hypothetical IMSI Catcher Overuse in Canada," *Citizen Lab – Telecom Transparency Project // CIPPIC*, retrieved December 7, 2017, https://citizenlab.org/wp-content/uploads/2016/09/20160818-Report-Gone_Opaque.pdf; Dalek, Jakub; Kleemola, Katie; Senft, Adam; Parsons, Christopher; Hilts, Andrew; McKune, Sarah; Ng, Jason Q., Crete-Nishihata, Masashi, Scott-Railton, John; and Deibert, Ron. (2015). "A Chatty Squirrel: Privacy and Security Issues with UC Browser," *The Citizen Lab*, retrieved December 7, 2017, https://citizenlab.org/2015/05/a-chatty-squirrel-privacy-and-security-issues-with-uc-browser/.

employees and paying $100 plus tax, a one-month sample of the cell towers (and the towers' locations) to which their mobile phone had connected while making calls.

Metadata about web server and website access is sensitive and attracts privacy protections under Canadian law.[38] Participants who filed DARs asked for access to the IP addresses historically assigned to end-users, or URLs of domains visited. This request was meant to better understand if IP addresses were retained beyond technical necessity by telecommunications providers. While most responding companies provided instructions on how the requester could obtain their currently-assigned IP address, historical customer IP addresses were challenging for many responding companies to dredge up from their systems. On this basis, they routinely declined to provide the data to requesters. Other companies revealed that they did retain the IP addresses assigned to mobile devices for a short period of time.

Yet others noted that their network configuration meant that mobile devices were never provided with publicly-addressable IP addresses and thus could be neither used to map users' activities nor provided to requesters. In still other situations, responses were incongruent with publicly-available information. For instance, Rogers Communications indicated it did not retain data concerning websites individuals visit, but analyses of how Rogers has modified web pages to warn wireline Internet customers that they are approaching their monthly bandwidth quota suggest that this type of information may be retained by the company's networking equipment for at least the duration required to insert these notices.[39]

Many of Canada's telecommunications providers also provide applications that let customers make modifications to their accounts or watch television. Past work has shown that applications installed on smartphones can collect extensive amounts of personal information,[40] so our DAR template includes questions about what information these mobile applications collect. Only providers operated by Rogers Communications provided meaningful responses to this question. They outlined the types of information collected and how some data could be deleted by removing the application.

4.1.2 HOW LONG INFORMATION IS RETAINED

Participants' DARs requested all records, which were intended to be used to infer how long different kinds of data were retained. In 2014 many telecommunications providers did not pro-

[38] R. v. Spencer, [2014] 2 S.C.R. 212.

[39] Brotherston, Lee. (2015). "mitm in telecom networks i told you so sort of," *SquareLemon Blog*, retrived December 7, 2017, `https://blog.squarelemon.com/2015/03/mitm-in-telecoms-networks-i-told-you-so-...-sort-of/`.

[40] Wandera. (2017). "Mobile Leak Report - 2017," *Wandera*, retrieved July 6, 2017, `http://go.wandera.com/rs/988-EGM-040/images/WP_MLR_v2.pdf`; Enck, William; Gilbert, Peter; Chun, Byung-Gon; Cox, Landon P.; Jung, Jaeyeon; McDaniel, Patrick; and Sheth, Anmol N. (2010). "TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones," *OSDI'10 Proceedings of the 9th USENIX conference on Operating systems design and implementation*, pp. 393-407.

vide public data retention schedules in response to public pressure for them[41] and therefore did not indicate to customers how long they retained what personal information. When providers did provide a clearer indication of retention periods their responses remained incomplete. Shaw, for example, indicated that their general retention period was seven years, but that there are also cases where data is retained for different periods of time (e.g. IP leases are retained for one year, whereas email mailbox information is stored for 60 days after closing an account).

Few companies explicitly stated how long they retained 'subscriber data', whatever its meaning. Providers generally stated that they retain billing information for seven years, so that a significant amount of subscriber information might be retained for an equivalent period of time, but not necessarily all of it. The same is true of calling and SMS/MMS information: metadata associated with these kinds of communication are retained; some companies asserted that content is never collected, that the metadata is disposed of after 13 months, or both. Even when companies did explain how long information was retained, the requester may not have been able to view the material without first paying a material fee. For instance, some data (e.g., calling information) was available through online billing records but Rogers, for instance, made only 18 months of data available to customers online. However, Rogers keeps records as long as seven years, suggesting that based on rates quoted, a long-time customer could have to pay up to $990 plus applicable taxes and charges to view this data.[42]

4.1.3 WITH WHOM INFORMATION IS SHARED

The earliest DARs issued by research participants in 2014 revealed that companies were uncertain as to whether they could tell customers that government agencies had requested their personal information. Some companies responded by stating that DARs would only be honored if the requester could obtain a court order. They then revised how they responded to requests: since any response could be contrasted with another person's response, requesters could compare their responses and in the case of variation realize that one of the person's records may have been disclosed to a government agency. Given the possibility of persons comparing their responses these companies considered themselves prohibited from confirming or denying that any information had been disclosed. This situation changed following a decision from the Office of the Privacy Commissioner of Canada, which established how companies are to respond to

41 Parsons, Christopher. (2014). "The Murky State of Canadian Telecommunications Surveillance," *The Citizen Lab*, retrieved July 6, 2017, `https://citizenlab.org/2014/03/murky-state-canadian-telecommunications-surveillance/`.

42 Rogers cited that it would cost $15/month to obtain any historical call records in excess of 18 months, which were accessible online. In the case of long-term customers which had seven years of calling records with Rogers, a requester would have to ask for 66 months of data to get a comprehensive record, assuming they could access the latest 18 months from their online account.

such questions.[43] In our 2016 data, the clearest TSP response to this question from Rogers stated that "[o]ur records reveal that no request/disclosure was made" for the relevant account.

Also in contrast with results from 2014, some telecommunications providers in 2016 informed post-paid customers that some information was shared with credit agencies each month. No other third parties were indicated by any telecommunications provider surveyed in 2016, although their own privacy policies and terms of service indicated that information may be shared with other third parties. Moreover, no mention was made of contracted parties having access to customer information, or such information being shared with contracted parties. This absence does not indicate that companies share data inappropriately but, instead, that responses provided by companies to this question remain threadbare.

4.1.4 AMBIGUITY OF RESPONSES

Past research has examined how privacy policies and terms of service, and some DAR responses, use ambiguous language when specifying the kinds of data collected, retained, or disclosed.[44] When analyzing the DAR responses from telecommunications companies we again found ambiguous responses: for instance, companies would "generally" not keep some types of data whereas they would retain other types. Some stated that "very little" information is retained when using a mobile application without identifying the specific kinds of information that are, in fact, retained. While using such language may reduce the risk that a company's response may be inaccurate, it also prevents customers from knowing exactly how, why, and for how long their personal information is collected and potentially disclosed to other parties–and may indicate an internal lack of clarity as to the precise nature of their practices. Even when companies did explain for how long personal information to which customers have access rights was retained, many participants were unable to view that information without first paying a fee. Companies often only provide customers with access to the previous 18 months of data through their online customer portals, despite often retaining some records for as long as 7 years. A long-time Rogers customer was asked to pay $15/month for data not included in the customer portal,

[43] Office of the Privacy Commissioner of Canada. (2016). "PIPEDA Report of Findings #2016-008: Investigation into a telecommunications company's response to an individual's request for access to information about disclosures of her personal information to other parties," *Office of the Privacy Commissioner of Canada*, retrieved July 6, 2017, `https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2016/pipeda-2016-008/`.

[44] Bennett, Colin; Parsons, Christopher; Molnar, Adam. (2014). "Real and Substantial Connections: Enforcing Canadian Privacy Laws Against American Social Networking Companies," *Journal of Law, Information & Science*; Bennett, Colin; Parsons, Christopher; Molnar, Adam. (2014). "Forgetting and the right to be forgotten." In Serge Gutwirth et al. (Eds.). *Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges*, Springer; and Shamess, Brittany; Smith, Michael; and Parsons, Christopher. (2013). "Long Summaries of Social Networking Service Privacy Policies," *The CATSMI Project*, retrieved December 7, 2017, `https://christopher-parsons.com/wp-content/uploads/2017/06/Long-Responses-1.0.pdf`.

which would have cost $990 plus applicable taxes and charges to obtain all of the information retained by Rogers. In most other cases, when participants received a response that indicated a fee would be required, the participants did not follow up with companies, thus terminating their requests.

Similarly, material fees were linked with accessing historical IP addresses associated with devices (in one case, a company informed a customer it would take 60 hours of labour, billed at $20/hour, to produce all of their IP address logs) or copies of SMS/MMS communications. This data can potentially greatly empower surveillance activities. [45] Absent ready access to this kind of information, telecommunications provider customers in Canada cannot ascertain whether similar activities are possible using data collected by their providers.

[45] As example, see: Parsons, Christopher; Israel, Tamir. (2016). "Gone Opaque? An Analysis of Hypothetical IMSI Catcher Overuse in Canada," *Citizen Lab – Telecom Transparency Project // CIPPIC*, retrieved December 7, 2017, `https://citizenlab.org/wp-content/uploads/2016/09/20160818-Report-Gone_Opaque.pdf`; Parsons, Christopher. (2016). "Transparency in Surveillance: Role of various intermediaries in facilitating state surveillance transparency," *Centre for Law and Democracy*, retrieved December 7, 2017, `http://responsible-tech.org/wp-content/uploads/2016/06/Parsons.pdf`.

Company	Request Date	First response	Notes on metadata access
Bumble	2016-06-21	2016-06-22	Bumble responded with a Subject Access Request form for participant to fill out, who did not do so.
Grindr	2016-06-15	2016-07-13	Grindr responded with some information about its privacy practices but did not provide any personal data, stated it required a subpoena to do so.[i]
OkCupid	2016-06-03	2016-07-01	OkCupid responded to some questions initially, but suggested participant could access personal data by logging into the service. After back and forth and identity verification, the requester received detailed data.
Tinder	2016-06-23	2016-07-19	Tinder responded to some questions initially, but suggested participant could access personal data by logging into the service. Tinder stated participant could get other data by verifying identity, which participant did not do.
Scruff	2016-06-15	N/A	N/A

[i] In December 2017, we received DAR correspondence between a requester and Grindr. In this correspondence, Grindr did provide data upon request. This correspondence is out of scope of this analysis, but warrants a mention to ensure Grindr is not misrepresented.

Table 4: Overview of online dating application DARs

4.2 ONLINE DATING APPLICATIONS

Four out of six online dating companies responded to DARs: OKCupid, Tinder, Bumble, and Grindr. OKCupid and Tinder are both owned by Match Group. Due to identity verification and jurisdictional barriers, only our OkCupid participant was able to get detailed data from the company.

4.2.1 WHAT INFORMATION IS COLLECTED

Table 4 presents an overview of the information provided by online dating companies in 2016. A full listing of the questions that were sent to each telecommunications company is available in Appendix A.

SUBSCRIBER DATA

Previous research that analyzed social media companies' lawful enforcement guidebooks noted that companies often include different kinds of information in the category of 'subscriber data'.[46] Some companies interpreted "subscriber data" to refer only to a customer name and address,

[46] Bennett, Colin; Parsons, Christopher; Molnar, Adam. (2014). "Forgetting and the right to be forgotten." In Serge Gutwirth et al. (Eds.). *Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges*, Springer; see also: The CATSMI Project Resources at `https://christopher-parsons.com/catsmi-project/`.

while others interpreted the term to include financial information, login locations, or other data elements.

DAR responses from dating application companies indicated different interpretations of the term "subscriber data," or whether it is a meaningful or relevant term altogether. Tinder declined to respond on the basis that subscriber information was available to users to view within the Tinder mobile application and thus the request was out scope of PIPEDA. OKCupid responded similarly but, after probing from the DAR requester, the company asserted that "basic info" included a range of data elements, including name, userID, email address, join date, login count, date of birth and gender, sexual orientation, age, education, location, account level, and reasons for deleting the account. Less detail was provided by Grindr, which included the participant's email address, subscription purchase information, and information associated with the user's public profile.

METADATA

Metadata is often more useful than content for drawing inferences about individuals' online activities, so research participants' DARs asked about IP addresses and device identifiers. Only one company, OKCupid, provided a dataset which indicated that it retained IP addresses, and only after the participant sought clarification. Other companies declined to provide information or simply indicated that the information was publicly-available and thus fell outside the scope of a PIPEDA-based request for personal information. The companies also displayed differences as to how they manage location information. Grindr only retained location information associated with the most recent login. OkCupid retained mobile GPS information collected at the start of each usage session, apparently indefinitely. Tinder did not state explicitly whether or not it retained location information, despite using such information for establishing matches between users of the application.

Users typically upload content to their profiles when signing up for dating applications. In some cases this process involves authorizing a company's application to take information from Facebook or another social media account. In other cases, the user manually inputs information, such as a profile or additional photos. With regard to data retention, all responsive companies indicated that they could retain chat logs, photos, or other user-uploaded content indefinitely. However, when such content was available online for logged-in users, Tinder and OkCupid asserted that PIPEDA guarantees a right of access and not a right of data provision; users were sometimes responsible for looking through the content stored in the application to determine everything held by the company in question. However, this puts the onus on the individual to know all the places where content might be located when companies do not provide a specific data inventory to requesters.

4.2.2 HOW LONG INFORMATION IS RETAINED

Online dating providers did not generally state explicitly how long personal information was retained following collection, nor what happened with this data after a user ceased to use their account. Grindr did assert that it keeps only the most recent GPS location linked with the last login. OKCupid did affirm that it retains all of the login IP addresses and GPS information collected when a user opened the mobile application. But there was no indication of when such information would actually be deleted by the company, if ever. Photos that a participant had uploaded to OKCupid and later deleted were included in the company's response, indicating that such data were retained after a user deleted it.

4.2.3 WITH WHOM INFORMATION IS SHARED

Information collected by dating applications can include health information such as HIV status, a type of information that routinely receives exceptionally high degrees of privacy protection, so DAR templates provided to research participants asked whether data had been provided to other parties in aggregated or individuated data sets. Only the Match Group services, OKCupid and Tinder, directly responded to these questions. They indicated to which third parties they had provided individuated data sets, but not aggregated data sets, on the basis that aggregated data fell outside the scope of a PIPEDA-based request:The responsive services did not explain what, specifically, was aggregated or the processes by which information was de-identified, preventing requesters from confirming the companies' assessments of the data's de-identification.

Tinder, OKCupid, and Grindr provided specific information concerning how they would respond to government requests for a user's data. Grindr would only release such information after being served with a "subpoena." Tinder and OKCupid explained that Canadian government agencies would need to use the mutual legal assistance treaty or letters rogatory process to obtain data from the company.[47] This explanation was made somewhat murkier, however, by the companies noting they would respond to law enforcement requests in a wide range of additional situations, including but not limited to efforts to investigate, prevent, or take action regarding illegal activity, comply with applicable laws or cooperate with law enforcement, or to enforce the services' policies or defend and protect their rights. This broad series of exceptions limit the usefulness of their general response as to the conditions under which they will disclose users' data.

[47] See Government of Canada. *Requesting Mutual Legal Assistance from Canada*. *http://www.justice.gc.ca/eng/cj-jp/emla-eej/mlaguide-guideej.html*.

4.2.4 AMBIGUITY OF RESPONSES

As with other kinds of industries, dating application companies often use words such as "could," "may," or "sometimes" to avoid blanket commitments as to whether data will be collected, how long they might be retained, or whether they will be disclosed to other parties. Such wording may reduce inaccuracy risk or retain latitude as to the range of lawful activity organizations can take with users' data, but prevents those whose data is being collected from knowing what is being collected, for what purposes, for how long, and to whom it is shared.

Such ambiguity carried over to how some companies discussed their security practices. Of the companies that answered how they secured user information, only two of the four were responsive. While Match Group companies said they took "appropriate security measures" they also warned that users "should take care with how you handle and disclose your personal information." The result is that users "should not expect [...] that your personal information, searches, or other communications will always remain secure." So while the companies maintain that they take appropriate measures to safeguard the sensitive information that is entrusted to them, users were effectively being told that they should not trust that those measures are necessarily sufficient to actually *protect* that information. The result is that a user may be uncertain about how much they should actually say or do on the application, or how they should evaluate the actual likelihood that their activities will be exposed to unauthorized third-parties.

4.2.5 COST OF DATA

None of the responsive companies charged a fee to provide users their data. None withheld data pending a fee.

Company	Request Date	First response	Notes on metadata access
Apple	2015-11-2	2015-11-13	Provided password-protected multi-sheet spreadsheet after verifying requester's identity over email.
Basis	2015-10-23	2015-11-23	Responded to some questions, said it would provide detailed data at later date but never did. Requester did not follow up.
Bellabeat	2015-12-15	2015-12-17	Responded to some questions, said it would provide detailed data at later date but never did. Requester did not follow up.
Fitbit	2015-10-22	2015-11-14	Responded to questions and provided fitness data spreadsheet after verifying requester's identity over email.
Garmin	2015-11-16	N/A	No response
Jawbone	2015-11-03	2015-12-08	Responded to some questions and directed requester to company data export tool.
Mio	2015-11-10	N/A	No response
Withings	2016-06-21	2016-08-23	Responded to some questions and directed requester to data export tool which included fitness data; gave requester option to manually request data via postal mail.
Xiaomi	2015-11-10	2016-01-12	Response did not address the DAR in any substantive way.

Table 5: Overview of fitness tracker DARs

4.3 FITNESS APPLICATIONS

Six out of nine fitness tracking applications in our sample responded to DARs, and provided detailed fitness information. The data download tools offered by several companies provide a convenient and relatively secure method for accessing fitness data. These data download tools do not provide detailed subscriber or login session information, and are not a substitute for full access to personal data.

4.3.1 WHAT INFORMATION IS COLLECTED

Table 5 presents an overview of the information provided by fitness tracking companies in 2015. A full listing of the questions sent to each telecommunications company is available in Appendix I; questions asked of fitness tracking companies were similar to those asked of telecommunications companies, adding further questions specific to fitness trackers.

SUBSCRIBER DATA

While fitness tracking companies do not typically use the "subscriber information" terminology in their privacy policies, we found that information provided at registration generally fell under this category. All the fitness trackers and apps in our sample collected users' names and email addresses upon registration. Fitness tracking companies typically offered their applica-

tions for free and did not verify identities beyond confirming email addresses, which would let privacy-conscious users input false names and throwaway email addresses. However, the applications also collected date of birth, gender, height, and weight which, if one wishes accurate fitness tracking, are less amenable to pseudonymity. This information is potentially sensitive if combined with other identifiers such as IP address or location.

Fitness tracking companies responded in varying ways after receiving DARs from our participants:

- Four of nine companies (44%) provided reasonably fulsome responses to the DARs.
- Three of nine companies (33%) – Mio, Xiaomi, and Garmin – did not provide responses to the substance of the DARs.
- Two companies (22%) – Basis and Bellabeat – responded to some of the questions contained in the DARs, and indicated that they would provide detailed data in a later communication, but failed to do so.

As to subscriber data retained, three of four responsive companies – Apple, Jawbone, and Fitbit – returned basic personal information such as height, weight, date of birth, and in the case of Apple, full name and mailing address. The fourth company that responded to the request, Withings, simply directed our participant to their data export tool. The data exported from that tool did not contain any subscriber information, and was instead limited to fitness activity logs.

METADATA

Metadata associated with fitness tracking activities can potentially reveal much more than basic records themselves.[48] For instance, data about a run might be limited to the time the run began, distance covered, and overall duration of the activity. On the other hand, when this run information is associated with geolocation or IP address data, it can become much more personally identifiable: such information can be particularly sensitive if the user refrained from providing their real name or address during account registration on the basis that it could be used to identify them. Our DARs asked fitness tracking companies for records of the IP addresses they had logged, as well as geolocation records.

Companies that responded to the DARs provided fitness data through one of two methods. First, Jawbone and Withings directed the requester to the company's data export tool. In these cases, participants authenticated with the fitness tracking services, and pressed a button to download a spreadsheet of their activity. Second, Fitbit presented two options for the secure transmission of data to the participant. The first was a shared Google Drive spreadsheet, and the second was an encrypted Zip file. The participant chose the Google Drive option.

[48] Scott-Railton, John. (2018). Fit Leaking: When a Fitbit Blows your Cover. Retrieved 9 February, 2018, `http://www.johnscottrailton.com/fit-leaking/`

Apple provided many instances where IP addresses were logged,[49] but none were associated with fitness activities (Apple stated it does not collect fitness data). Basis claimed it "does not track which IP addresses are used." Fitbit provided a participant with a dataset of over 25,000 timestamped IP address logs – an average of 157 per day. Withings did not explicitly address the request for IP addresses and geolocation information but did link the requester to a privacy policy stating that Withings applications collect "Cookies & Technical Features." Withings did not explicitly define these categories in its policy, but they may potentially include IP addresses.

When asked to provide access to personal fitness data they collect, Jawbone, Withings, and Fitbit provided spreadsheets with step counts, calories burned, weight, distance covered, and other data, organized by day. Jawbone's spreadsheet additionally included columns for mood, gender (binary), and mealtimes. Withings' spreadsheet also included columns for elevation, blood pressure, and oximetry. Fitbit's also included minutes spent active per day. Fitbit included a further spreadsheet for heart rate data: over 18,000 timestamped heart rate records, taken at five-minute intervals over the six-month period the requester was using the device.

4.3.2 HOW LONG INFORMATION IS RETAINED

We were able to infer what appear to be the retention periods for some types of data for companies (Fitbit, Jawbone, and Withings) that prepared data disclosures of fitness information. For example, in its data dump, Fitbit included the date on which the device was first synced with Fitbit's servers. This timestamp corresponded to the date in the first entry of step records, as well as the first entry in IP address records. The earliest records corresponded to the date, nine months prior, at which the requester first synced their device. This suggests that Fitbit's IP address and step retention period is at least nine months, whether or not some cut-off exists. While Fitbit users would perhaps expect that their step and heart rate data would be retained indefinitely so that they can track their fitness over time, it is less clear that similar expectations would be in place for IP address data. Jawbone and Withings similarly retained fitness data, though unlike Fitbit, they did not provide access to historical IP address records.

Company positions on sharing information with third parties varied:

- Apple explicitly stated that it "does not share personal information with insurance companies, in an aggregated form or otherwise." Basis also explicitly stated that it had not provided the requester's data to insurance companies.
- Bellabeat responded more generally that data collected from their service "have not been provided to any 3rd parties nor will it be," and that they would ask permission from users before sharing any data.

[49] Apple's DAR response included personal data it had collected on the participant across many of its different services.

- Jawbone said it does not "rent, sell, or otherwise share your personal information," but included many exceptions, such as "for the purposes of a business deal (or negotiation of a business deal) involving sale or transfer of all or part of our business or assets." Fitbit said it does not provide identifiable information to third parties outside of the purposes identified in their privacy policy, but would provide aggregated, de-identified data to partners.

All responsive companies indicated they would respond to valid lawful requests for access to their users' personal data. Two topics where variation emerged were on the subject of which jurisdiction could serve these legal requests, and whether or not the company would endeavour to notify users that such requests had occurred.

Regarding the question of jurisdiction, Apple and Fitbit stated they would only respond to U.S.-issued requests, with the remaining three responsive companies defining no explicit criteria for which country could serve them requests. Apple directed our participant to its policy on government information requests, which indicated that the company requires a search warrant for all U.S. requests, and that international requests for U.S.-hosted data must comply with the U.S. Electronic Communications Privacy Act (ECPA). Fitbit similarly stated it would only comply with a "valid legal process issued by a U.S. Governmental entity or court and when properly served," which included international requests processed by U.S. authorities under Mutual Legal Assistance Treaties (MLATs). In contrast, Basis stated its "policy is to comply with applicable laws and regulations in the jurisdictions in which it does business." Jawbone and Withings had similar responses to Basis regarding legal compliance.

Regarding whether or not the company would notify users in the event of a lawful request for access to their personal data, Apple and Withings made statements, while the remaining four companies did not mention notification at all. Withings asserted it would notify users of any disclosures "apart from where this is prohibited." In a slight variation, Apple indicated specifically that it would give prior notice to customers about disclosures to law enforcement and other governmental agencies, if not prohibited by law.

4.3.3 AMBIGUITY OF RESPONSES

Some DAR responses were inconsistent with published policies.

When companies were asked about whether or not a legal dispute with the Canadian requester would be held to non-Canadian laws, Basis stated that "Where issues can and should be raised/resolved will depend on, among other things, the specific nature of the issue and the parties involved" – but Basis' Terms and Conditions explicitly states that both parties "consent to the personal jurisdiction of the state and federal courts of Delaware." [50]

[50] See: `https://www.mybasis.com/legal/tos/`. Whether or not Canadian courts will give effect to a choice

In the same way, Bellabeat wrote in their response that personal data transfers between their device, their app, and their servers were "secured against all potential attacks," but their privacy policy states they "cannot guarantee the security of personal data during its transmission or its storage on our systems."

This discrepancy no doubt has much to do with avoiding over-broad claims that would attract a high degree of company liability. From a consumer standpoint, however, it is also another example that highlights the importance of companies ensuring that, if consumers are to treat a company's statements as accurate and representative, their DAR responses ought to align with their stated policies, and that both reflect the company's actual practices. Otherwise, consumers will be left in the dark about what actually happens to their personal information– the opposite of the privacy rights which they are guaranteed.

4.3.4 COST OF DATA

No responsive company charged a fee to provide users their data. None withheld data pending a fee.

of foreign jurisdiction in a service provider's contract of adhesion with a consumer is a more fraught matter: see, e.g., Douez v. Facebook, Inc.,*2017 SCC 33*. Here, our focus is rather on the inconsistency between the company's own statements.

5 COMPARISON OF RESPONSES ACROSS INDUSTRIES

This section discusses the variations between different industries with regard to information collected, data retention times, with whom data is shared, ambiguity or clarity of responses, and the costs of access.

5.1 INFORMATION COLLECTED

Requesters' DAR letters asked questions about their providers' collection of their data, metadata, geolocation data, and mobile application data. We observed significant variation in the responses from companies both within and across industries.

5.1.1 SUBSCRIBER DATA

There was significant variation within and across industries regarding how the companies defined and discussed "subscriber data". In some cases subscriber data included very little information: name, physical address, and email address. In other cases it included all the information included in public or semi-public profiles, payment information, and metadata generated when creating the account.

Telecommunications companies were most resistant to providing information about the metadata they collected in the course of offering consumer services. In some cases they made it available only contingent on paying hundreds of dollars. This resistance was based in part on the difficulty in retrieving the information, but meant that requesters were largely left in the dark about what specific data was collected. Some fitness tracking and online dating services, on the other hand, provided large data files of the collected content and metadata that users had generated in the course of using the respective companies' services. In doing so, however, companies often failed to respond to the specific questions raised by the requesters. In our view, the contrast likely arises because telecommunications companies operate a mix of physical infrastructure and software-enabled service offerings on the basis of systems developed incrementally over years. Their data operations may be more distributed across less-than-compatible packages, and more embedded in legacy systems, than online dating and fitness tracking applications. The latter applications, by comparison are relatively newer. They have been subject to far less built-up incrementalism, suggesting their systems may be more centralized. They are more novel applications, and are therefore less likely to rely on less-self-managed third-party packages than telecom operations, for which a large range of third-party billing and service delivery solutions exist.

5.1.2 GEOLOCATION DATA

Geolocation data was inconsistently provided by companies. There was no consistency in the kinds of location data, or the amount retained, by online dating or fitness tracking services. And telecommunications providers would only provide location data linked to cell towers at a fee, though some also noted whether or not such information had been remotely collected via the e911 system for either emergency response purposes or law enforcement surveillance.

5.2 METADATA

Some companies provided only limited amounts of information about what information they collected from their users. While some information might be provided, users would in many cases be instructed to just open their accounts to see other information held by the organization. This response was justified by some on the basis that PIPEDA functions as a right of access as opposed to a right of data provision.

5.3 HOW LONG IS INFORMATION RETAINED

The DAR letters that individuals sent to companies did not specifically ask for data retention schedules. However, we hypothesized that asking companies to provide requesters with copies of all the data they possessed would functionally result in companies revealing their retention periods. This hypothesis generally did not hold.

Some telecommunications providers revealed how long certain information, such as billing records or metadata pertaining to SMS messages, were retained but did not provide comprehensive retention schedules. For instance, information about retention periods of email accounts linked with their telecommunications services or information retained by cellular towers about their location were not discussed.

Fitness tracking and online dating services were rarely direct in explaining how long they retained information, though when they provided data dumps it was possible for requesters to infer for how long different kinds of information might be retained. However, these data dumps themselves were not always easily understandable: when supplied as large spreadsheets, it was left to recipients to ascertain what the categories referred to. Nor were these dumps likely comprehensive. These dumps were, however, provided free of charge and enabled requesters to learn more about the data collected about them, in stark contrast to telecommunications providers' failure to provide equivalent kinds of information.

5.4 WITH WHOM INFORMATION IS SHARED

There was significant variation in how industry sectors responded to questions regarding third-party sharing with commercial and government agencies.

Some telecommunications providers disclosed more information about the commercial entities with which data was shared (e.g., credit bureaus), though not all provided even this level of detail. Fitness tracking and online dating companies were more willing to identify whether information still linked to identifiable persons would be shared and, in some instances, the kinds of companies with whom it would be shared. However, when it came to the sharing of 'de-identified' or 'anonymized' data, respondents tended not to identify with whom the data was shared, sometimes on the basis that such information is not personal information. Without knowing what data is being shared and with whom, it is impossible for requesters to know the extent to which data was actually de-identified.

In 2016, Canadian telecommunications providers provided specific information about whether information had been shared with law enforcement agencies, which for most was a significant change from more ambiguous statements provided in 2014. Some online dating companies and fitness trackers provided information about when they would release information – such as with specific kinds of court orders – but more generally failed to specify the legal standards that would have to be met before they would share information. Perhaps most disturbing were situations where companies provided very specific requirements that had to be met before they would share personal information with government agencies, only to then include such a broad range of exceptions as to make the specific requirements effectively meaningless.

5.5 AMBIGUITY OF RESPONSES

The responses to DARs revealed that organizations tended to either provide highly couched responses, which makes them ambiguous, or provide contradictory statements or information. Some organizations would make strong assertions concerning data privacy when responding to a DAR that stood in contradiction to their formally stated privacy policies, leaving individuals who had read both without a clear understanding of a company's position on either the degree to which individuals' data was secure or where complaints had to be litigated. Moreover, companies across industry categories often used qualifying statements such as "may", "could", and "sometimes" when explaining how they treated, handled, collected, or shared personal information. Such qualifying statements limit the explanatory nature of DAR responses and, as such, reduced the utility of responses to the persons who receive them.

5.6 COST OF DATA

Apart from most telecommunications providers, no responsive organization charged a fee for providing access to collected data. PIPEDA principle 4.9.4 states that access should be provided at no or minimal cost.[51] However, telecommunications companies demanded hundreds of dollars before providing metadata information. Such costs disincentivize Canadians from learning about what information is collected and impair their ability to review data to understand what is collected and whether inaccuracies have been recorded.

[51] Office of the Privacy Commissioner of Canada (2013). "Interpretations Bulletin: Access to Personal Information," *Office of the Privacy Commissioner of Canada*, retrieved October 17, 2017,. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda-compliance-help/pipeda-interpretation-bulletins/interpretations_05_access/.

6 PARTICIPANT REFLECTIONS ON COMPANY RESPONSES

As part of this project, we conducted a small-scale survey of how Canadian respondents felt about the responses provided by companies to which they had issued data access requests. Findings from this survey included:

- Most AMI users were unfamiliar with their right to access their personal data
- Most requesters got a response from companies
- Inconsistent levels of satisfaction with how companies respond
- Most requesters found parts of the data they got back hard to understand

6.1 PARTICIPANT EXPERIENCES

Our survey was sent to a total of 197 individuals who had previously chosen to be contacted by the research team when they generated their data access request on AMI. Out of 197 individuals, 19 consented to take part in the study. Our focus for choosing participants was based purely on diversifying the pool of companies to which participants had made requests. We were not focused on determining differences in responses between genders, ethnic groups, or different age groupings and we did not collect this level of information.

The small number of participants make the results a qualitative sampling of participant experience that does not generalize to the broader population of AMI users (over 6000 as of February 2018).

Our surveyed participants made their requests between June 2016 and October 2016. They had previously learned about companies' privacy practices from a range of sources: corporate privacy policies (31.6%), social media (15.8%), customer service (15.8%), and "other" (31.6%).

- Three (15.8%) of the requesters were very familiar with their right of access to personal information using data access requests.
- Twelve (63.2%) were either very unfamiliar, or somewhat unfamiliar with these rights (each 31.6%).
- Four users (21.1%) were moderately familiar with their rights.

All of the participants stated that they were interested in learning what information was collected about them and how long such data was retained. Some expressed interest in specific data (e.g. metadata, GPS logs, browsing history) others had more general concerns (e.g. in the "data they collect and store about me"). One participant was interested in learning how seriously the company from which they were requesting data would take the request, whether the

data would indicate any relationship between the company and state agencies, and about the kinds of data being retained about their activities.

One of our research questions asked: "what proportion of organizations contacted would respond in any way to individuals' requests for access to their personal information?" Only twelve of the nineteen requesters received a response, whereas the remaining seven DARs were met with silence. All the responsive companies contacted requesters within thirty days. Four of the requesters who received a response were told that the responding company had availed itself of the time-limit extension that PIPEDA permits when the initial 30-day period would unreasonably interfere with ordinary activities, prohibit necessary consultations, or prevent conversion into an alternative format. [52] Four of the participants indicated that the response they received to their DAR did not include the requested data, and the remaining eight indicated they did receive their data.

PIPEDA requires that organizations respond to data access requests at minimal or no cost to the individual,[53] "imply[ing] that any fee charged should be a token one"[54] even where such a fee would not recover the data steward's costs.[55] All responsive companies returned data without charging a fee, but some cases companies made more comprehensive data disclosures contingent on payment. In one case, Rogers Communications provided a participant with geolocation data and voice/SMS metadata records in exchange for a fee: after significant email and telephone back-and-forth, these were established at $100 for preparing a month of voice/SMS metadata records, and $100 for cell tower geolocation data (totalling $226 inclusive of tax) based on $100 per hour (plus sales tax) for labour. The participant mailed a personal cheque to Rogers' legal department in order to receive the records, which were sent over email.

Participants disagreed about how well the companies' respective responses helped them understand how their data was collected and used. In four cases (33%) participants found the provided data somewhat or very helpful, three (25%) found it somewhat or very unhelpful, and the remainder were neutral (42%). In comments provided to us, one of the participants noted "none" of their data was missing from the response from the company they contacted. In an-

52 PIPEDA, subsection 8(4).

53 PIPEDA, Schedule 1, clause 4.9.4.

54 Office of the Privacy Commissioner of Canada. (2007). "Fees for access questioned, PIPEDA Case Summary 2006-354," *Office of the Privacy Commissioner of Canada*, retrieved December 7, 2017, https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2006/pipeda-2006-354/.

55 Office of the Privacy Commissioner of Canada. (2002). "Individual denied access to personal information, PIPEDA Case Summary 2002-111," *Office of the Privacy Commissioner of Canada*, retrieved December 7, 2017, https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2002/pipeda-2002-111/; Officce of the Privacy Commissioner of Canada. (2004). "Company refuses former employee's request for access, PIPEDA Case Summary 2004-285," *Office of the Privacy Commissioner of Canada*, retrieved December 7, 2017, https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2004/pipeda-2004-285/.

other case, a participant asserted that their service provider told them that "any information [the company] do get from my [device] is not easily read and in most cases is useless to anyone."

We also asked participants to explain in their own words how well they understood the responses companies had provided. They wrote that they had difficulty understanding the responses or disagreed that the responses were comprehensive. For one participant, the answers were "complicated to understand." Another found the response language somewhat legal and technical, and "somewhat difficult to understand." The remaining participants who provided commentary raised concerns with companies' responses: the explanation of whether information was disclosed to third parties was unclear, the response to jurisdictional questions not comprehensive, the companies did not provide fulsome details concerning how long data was retained; and companies generally refused to disclose whether they had shared or exposed information to third parties, including state agencies.

6.2 ANALYSIS OF PARTICIPANT COMMENTS

Participant responses highlight challenges facing companies that receive data access requests generated using the AMI web application. A significant proportion of the 19 participants who responded to our survey (36.8%) received no response. And of those whose DARs were met with some kind of response, 33% of participants stated the companies failed to provide them with their data. A basic means of improving requester satisfaction, then, would entail responding to requests and completely providing requested data.

As noted above, a number of participants found it challenging to understand the responses provided to them. In our examination of responses, we found that not all companies provided specific responses to each question, instead providing template letters that required requesters to examine the questions they asked and independently determine if their questions were responded to at all. Furthermore, some of the participants were disappointed that requests that should have elicited information about the types of data collected, retention periods, and parties to whom information was disclosed were not met with fulsome answers. Companies could overcome at least some of these limitations without increasing case-by-case workload by publishing comprehensive data retention schedules–something they are in any case required to maintain internally by PIPEDA and its counterparts. [56] Moreover, given that most of our participants' requests were motivated by a paucity of information as to what data was collected, retained, or disclosed, many such requests could no doubt be avoided altogether by publishing

[56] PIPEDA, Schedule 1, clause 4.5.3 ("Organizations shall develop guidelines and implement procedures to govern the destruction of personal information"); Insurance provider revises retention period and practices for insurance quotes containing personal information, PIPEDA Report of Findings 2014-019, 30 October 2014 ("Lessons Learned").

retention schedules and the terms under which retained personal information will be disclosed to third parties like government agencies.

Large telecommunications companies in Canada are notorious for issues with customer satisfaction. Lack of consumer trust in Canadian telecoms may, in turn, instil doubt in a provider's responses.[57] When told that a particular kind of data is not collected or retained, for instance, a customer may choose to disbelieve that the provider fails to collect that kind of data rather than take the statement at face value. How and whether to alter subscriber perceptions of an industry or its member companies is beyond the scope of this report. But companies can mitigate some of these challenges by working with third parties–including the developers of the authors' AMI web application. After one company explained to the us how its subscribers disbelieved the responses the company was providing, our team added language to that application to provide for scenarios where the provider does not collect or retain the information an end-user seeks.

[57] See, e.g., Goldberg, Mark. (2010). "We love to hate our service providers," *Telecom Trends: A Canadian Perspective on Trends in Telecom (blog)*, retrieved December 7, 2017, `http://mhgoldberg.com/blog/?p=3998`; Krashinsky, Susan. (2015). "Telus knows you hate telcos", *Globe and Mail*, retrieved December 7, 2017, `https://www.theglobeandmail.com/report-on-business/industry-news/marketing/telus-knows-you-hate-telcos/article24433980/`.

7 LIMITATIONS OF DATA ACCESS REQUEST PRACTICES

Data access requests do not, alone, deliver comprehensive answers to customers' questions about the personal information their service providers collect as a byproduct of using the service providers' services. In some cases, responses may be lacking because a company declines to indicate they received the request; in others, because companies believe neither that they are required to respond fully, nor that they ought to. These limitations are sketched out in additional detail below, along with some ways of overcoming them.

7.1 NON-RESPONSIVE COMPANIES

Companies may fail to respond to a DAR for a variety of reasons. They may lack business processes to act on the letter. They may lack internal resources to respond. They may not believe that they are obligated to respond, nor ought to.

Some businesses, especially smaller organizations, may not respond to a DAR if no one is designated to take action on such requests. Even when there is a designated person to receive the letter they may not be trained in how to respond and, rather than request clarification or learn the obligations they are under to respond, fail to provide any kind of response. Further, even where a business has designated someone responsible for receiving customer information requests, including DAR letters, who knows how to respond to them, the designate may be on the wrong side of a disconnect between policy or business operations matters, on one hand, and technical, on the other hand: a standard AMI DAR's question pertain to both. In these cases, a gentle reminder reiterating the reasons and rationales for each question, providing helpful guidance on how to proceed with the request, and noting the implications of not responding to the request,[58] may help to encourage the company to respond to the DAR. Companies may not themselves have yet developed a retention schedule or listing of all of the personal information that they retain. DARs and reminders may prompt companies to develop better data inventory and management procedures so that they can respond to their customers' DARs more efficiently and more accurately.

Some organizations believe that a DAR does not trigger a disclosure obligation, either because they dispute PIPEDA's or its counterpart's jurisdiction over the relevant activity, or because they believe that disclosure has already been provided. This reasoning, where relied on, may not be clearly communicated to the requester. In these cases, a follow-up letter that clearly asserts that the cited legislation and consequences for failing to respond may create clarity.

Should companies fail to respond to a DAR letter at all, even following reminders, requesters can notify their relevant privacy or data protection commissioner to initiate a complaint about

[58] PIPEDA, sections 14-16 (Federal Court) and subsection 20(2) (name-and-shame powers).

the company's behaviour. They might also contact members of the local media who may have a view as to whether there is a public interest in the company's non-response. Allocating resources to looking into the matter may be a reasonable corporate response both to negative press attention and to a Privacy Commissioner investigation. These resources are not without cost.[59]

Companies may also be motivated to improve their DAR responses and related transparency practices through competitive self-interest. Were DAR practices regularly ranked between competitors within an industry segment in a standardized manner, published on the web, and the rankings received a significant amount of attention, then companies might want to appear more highly ranked than their competitors. Initiatives using this approach include the Electronic Frontier Foundation's "Who Has Your Back?" annual look at tech company practices on privacy, security and freedom of speech,[60] annual look at tech company practices on privacy, security and freedom of speech, and the IXMaps ranking of Canadian telecommunications companies on various privacy practices."[61]

7.2 DISAGREEMENT AS TO WHETHER PIPEDA APPLIES

Companies sometimes assert, in writing, that they do not have to respond to a DAR letter which cites PIPEDA on the basis that they do not believe PIPEDA applies either to their organization or to the kinds of data that are being requested. In others, companies may assert that they disagree whether PIPEDA applies but nonetheless provide information pursuant to either their European data protection operations or as a goodwill gesture.

If companies insist that PIPEDA, as legislation, does not apply to them, it is potentially helpful to explain how companies with significant commercial connections to Canada are obliged to adhere to the PIPEDA principles and respond to requests made by consumers pursuant to PIPEDA. It can be helpful to cite legal cases where the Office of the Privacy Commissioner or Canadian courts applied PIPEDA to foreign-based companies, and how those companies recognized they were legally obliged to modify their businesses practices as an outcome of the Commissioner's recommendations.[62] When a company persists in not responding to a request

[59] Office of the Privacy Commissioner of Canada. (2016). "Investigation into a telecommunications company's response to an individual's request for access to information about disclosures of her personal information to other parties, PIPEDA Report of Findings 2016-008," *Office of the Privacy Commissioner of Canada*, retrieved December 17, 2017, `https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2016/pipeda-2016-008/`.

[60] Electronic Frontier Foundation (2017). "Who has your back?". `https://www.eff.org/who-has-your-back-2017`.

[61] Clement, Andrew and Obar, Jonathan, A. "Keeping Internet Users in the Know or in the Dark: A Report on the Data Privacy Transparency of Canadian Internet Carriers". *IXmaps.ca & New Transparency Projects*. `https://www.ixmaps.ca/docs/DataPrivacyTransparencyofCanadianCarriers-2014.pdf`.

[62] E.g. Office of the Privacy Commissioner of Canada. (2009). "Report of the Findings into the Complaint filed by CIPPIC against Facebook Inc.," *Office of the Privacy Commissioner of Canada*, retrieved December 7, 2017„

because it does not believe PIPEDA applies the only solution may be filing a complaint with the Office of the Privacy Commissioner of Canada or contacting the media to see if, subsequently, the business recognizes binding commitments to Canadians under PIPEDA or its counterparts.

Businesses may recognize that they are obligated to respond to a request but disagree that PIPEDA obligates them to provide all that has been requested. There are at least two kinds of responses they might provide. First, a business might assert that some kinds of data, such as profile information, does not need to be disclosed because a customer can log into the relevant company's service offering and inspect the information for themselves. This kind of response is rooted in an understanding of PIPEDA as granting a right to access one's personal information,[63] but not to receive a copy of that information.[64]Put another way, some companies may only provide copies of information for a requester when the organizations believe that the information is otherwise inaccessible. Second, a business might recognize that some information need be disclosed under PIPEDA but assert that other information need not, such as certain metadata about telecommunications services or analytics information that is collected in the process of using the company's services.

In the first case, it can be helpful to ask a company to specify where all of the public information they assert they do not have to provide to you is located. When creating a new profile with a service a customer might not be fully aware of all the places information is stored and, as such, it is within a requester's right to ask for clarity about where it is stored in a company's service offerings. Furthermore, when an account was created on the requester's behalf – such as a child setting up their parents' new fitness tracker account – the subscriber to the service might be unaware of what information was input during the setup process. In the second case requesters can explain why they think that the relevant information should be disclosed. In the case of 'anonymized' information, as an example, a requester can explain that they need some copies of this information in order to confirm that the data has in fact been anonymized.

7.3 INCOMPLETE RESPONSES TO PIPEDA REQUESTS

The DARs sent by our participants (See Appendix I) pose specific questions concerning businesses' collection, processing, and handling of data access requests. These questions were structured so that companies could granularly respond to each, instead of responding to a slew

`https://www.priv.gc.ca/media/1033/2009_008_0716_e.pdf`; see also: Office of the Privacy Commissioner of Canada. (2010). "Reaching for the Cloud(s): Privacy Issues related to Cloud Computing," *Office of the Privacy Commissioner of Canada*, retrieved December 7, 2017, `https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2010/cc_201003/`.

63 PIPEDA, Schedule 1, clause 4.9 ("[u]pon request, an individual shall be informed of the existence, use, and disclosure of his or her personal information and shall be given access to that information").

64 See, however, PIPEDA, Schedule 1, sub-clause 4.9.4 ("[t]he requested information shall be provided or made available in a form that is generally understandable").

of questions using a single response. However, sometimes businesses simply fail to respond to certain questions; in other cases they do not affirmatively state whether information is available or not; and in others may provide a response that does not answer the specific question being asked. In several cases, process issues such as cost, identity verification, and security procedures precluded participants from receiving fulsome responses to their data.

While it may sometimes seem to a data subject as though a company has not responded to a particular question, the company's delegate may think otherwise. In these situations, simply asking a company to more clearly respond to particular questions can result in clearer answers, though sometimes a series of back and forths with a company is required before receiving a satisfactory answer.

In other cases, organizations may decline to respond to certain questions because the questions asked either do not pertain to the business's services or because there is no responsive data to provide. That does not, however, mean that the company will articulate explicitly that the non-response flows from the aforementioned reasons. Contacting companies and asking them to positively assert that either they do not provide service offerings of the nature described, or do not possess any responsive data to the question asked, can help better delineate a company's actual practices.

Companies are often willing to communicate with customers who have submitted DARs, because they want their customers to trust their services. As a result, companies may be willing to provide greater clarity when initial responses are unclear. When a company is not willing to engage in a discussion, however, its customers can remind it that failures to respond may lead to a complaint to the Office of the Privacy Commissioner of Canada.

Company processes presented hurdles for several of our participants in receiving fulsome responses to their DARs. Generally, these hurdles were associated with the cost of receiving a fulsome response, identity verification requirements, or other security procedures.

In the telecommunications industry, the most prevalent barrier to access was fees, whose payment was a condition to most companies' provision of detailed geolocation or SMS metadata; in the case of Fido, the minimum cost was $100, for an hour's labour.

Identity verification raised a challenge for participants sending DARs to online dating companies. For example, OKCupid would not provide copies of their retained data to our participant without the participant first mailing the company "a notarized copy of your driver's license or passport." The participant was able to negotiate with the company to email a photo of a redacted driver's license to verify their identity.

In the case of fitness tracking services, simple control over the email address used to register the requester's account was sufficient verification. The barrier in the fitness tracking industry were security mechanisms for the confidential transmission of data. In one instance, a participant was asked to provide the company with a PGP public key so that the company could send

encrypted data to them. PGP is a notoriously difficult-to-use technology that few Canadian consumers are likely to be familiar with.[65] In another instance, a participant had to provide the company with their personal Google account so that the company could share a document with them using Google Docs. In both these cases, access was predicated on the requester using a new technology that had not previously been needed to use the companies' services.

7.4 OVERCOMING DATA ACCESS REQUEST LIMITATIONS

Ultimately, DARs only provide as much information about a company's practices as the company chooses to reveal. In some cases, companies may provide very detailed explanations of their data-handling practices whereas, in others, they may only provide responses to a handful of questions or decline to respond to the questions at all. Furthermore, the information provided through a DAR may differ from how a company explains its activities in other venues, like terms of service, privacy policies, discussions with the media, or lawful access guidebooks or transparency reports.

There are a few ways of overcoming DAR limitations. First, DAR letters can be sent while simultaneously conducting policy or technical research. Policy research can include analyses of how companies assert they will, do, or may collect information from customers in their privacy policy documents or terms of service. If a company's online policies say one thing, and companies' DAR responses say another, then customers or researchers can subsequently inquire about which corporate statement is more accurate. Alternately, such research may involve publicly-accessible documents outlining how government agencies can gain access to customer information that is held by the business; such documents are often referred to as either government access guides or law enforcement handbooks. Such documents can explain what kinds of information are collected by the company and the terms under which the company will disclose data to government agencies. The effect of reading these documents is to provide insight into whether a company has been fully responsive to a DAR letter about the kinds of information the company can collect and link to specified persons. It is also possible that law enforcement handbooks can be obtained using access to information and privacy legislation, which citizens can use to compel government agencies to produce documents about their activities.

Technical research can also be conducted to evaluate how devices or software collect, handle, and transmit information. Tests can be conducted to determine the level of security used to protect personal information sent to the company, as well as whether there are significant vulnerabilities in how the company collects, handles, or stores personal information. Used in tandem with a DAR response, technical research can clarify whether best security standards

[65] Alma Whitten and J. D. Tygar. (1999). "Why Johnny can't encrypt: a usability evaluation of PGP 5.0." In *Proceedings of the 8th conference on USENIX Security Symposium*, Vol. 8. USENIX Association, Berkeley, CA, USA.

are used and whether the information provided in response to the letter correlates with actual business practices.[66]

Some unknowns, however, are still not easily addressed. As an example, it is unclear whether a DAR would entitle a customer to better understand how a company uses algorithms to learn more about a customer based on the data they have submitted to the company. And absent the algorithms working in a way that is open to independent research, it may be impossible to know the kinds of additional data that the algorithm may be producing. In aggregate, these challenges mean that while the baseline information that a customer presents to the company might be revealed through a DAR response, technical research, or policy analyses, the secondary creation of information about the individuals' personal information may be inaccessible using these methods alone, particularly if the algorithms' outputs are presented in real time but not retained. As suggested by the rise of algorithmic fairness as a area of concern beyond traditional information access,[67] insight into algorithmic bias may not be forthcoming from DAR responses, even supplemented with technical or policy research.

[66] See Hilts, Andrew; Parsons, Christopher ; and Knockel, Jeffrey. (2016). "Every Step You Fake: A Comparative Analysis of Fitness Tracker Privacy and Security," *The Citizen Lab and Open Effect*, retrieved December 4, 2017, `https://openeffect.ca/reports/Every_Step_You_Fake.pdf`.

[67] Rainie, Lee and Anderson, Janna. (2017). "Theme 7: The need grows for algorithmic literacy, transparency and oversight." *Pew Research Center: Internet and Technology*. Retrieved 9 February, 2018, `http://www.pewinternet.org/2017/02/08/theme-7-the-need-grows-for-algorithmic-literacy-transparency-and-oversight/`.

8 RECOMMENDATIONS TO IMPROVE DATA ACCESS REQUESTS

While DARs can help researchers and consumers better understand company data collection, retention, and disclosure practices, the process that unfolds between requesters and companies can be improved. Below we suggest several practices companies can undertake to expedite consumer inquiries, improve communications with customers, and improve DAR responses before requests are ever issued.

8.1 PRIOR TO REQUEST

8.1.1 RETENTION SCHEDULES

Our initial motivation for using DARs as a research tool was to better understand the types of data that certain companies collected and for how long they stored them. This information was not provided to the public in a clear manner. Similarly, consumers using the AMI web application often indicated that they were motivated to better understand how their data was handled by companies. It stands to reason, then, that the number of requests that companies receive should decline if companies publish their data collection and retention schedules. Particularly as internal retention guidelines are required by PIPEDA, public retention schedules would likely satisfy many of the questions held by researchers and consumers.

RECOMMENDATION

Companies should prepare and publish data retention schedules that identify the specific types of information they collect, and the period of time for which they retain the identified information. Companies could consider the model recommended in the DIY Transparency Reporting Tool published by Citizen Lab, which is based on an analysis of types of data that are often retained by telecommunications and digital-first companies.[68]

8.1.2 DISCLOSURE PROCESSES

Another motivation for using DARs as a research tool was to better understand whether information held by a company had been disclosed to other parties, including government agencies. Several consumers who used the AMI web application also indicated this question was important to them. Some international companies publicize the terms under which they provide information to government agencies in 'government access handbooks'.[69] While these hand-

[68] Parsons, Christopher. (2016). "Release: DIY Transparency Reporting Tool," *The Citizen Lab*, retrieved December 7, 2017, `https://citizenlab.org/2016/06/release-diy-transparency-report-tool/`.

[69] e.g. Apple. (2017). "Privacy - Government Information Requests," *Apple*, retrieved November 2, 2017, . https://www.apple.com/ca/privacy/government-information-requests/.

books do not indicate whether any particular person's information has been disclosed to a government agency – this is something individuals must ask in their DAR letters – they may assuage concerns that information is being inappropriately disclosed. Spelling out the specific rationales and processes government agencies must undertake before they disclose subscribers' data may improve trust between consumers and the companies to which they entrust their personal information, and thus reduce the likelihood of receiving DAR letters because individuals are comfortable with the terms and processes that a company has established for disclosing information with government agencies and corporate third parties.

RECOMMENDATION

Companies should prepare and publish government access handbooks that identify the different kinds of personal information held by the companies and establish the specific legal powers and processes that must be undertaken before the company in question will disclose any of its subscribers' personal information. Companies could consider the model recommended in the DIY Transparency Reporting Tool published by Citizen Lab, which is based on an analysis of the government access handbooks already published by some American companies.[70]

8.1.3 TRANSPARENCY REPORTS

To further improve the trust between consumers and companies, and to mitigate concerns consumers may have about personal information being disclosed in bulk to government agencies or through civil proceedings, companies could release transparency reports which disclose the aggregate number of times they have disclosed subscriber information to third parties. In past work we have discussed the importance of such reports for informing the public about corporate actions;[71] these reports will also notify the public about the regularity with which information is shared with other parties and may lead to a corresponding decrease in the regularity with which consumers file DAR letters in an effort to understand how often information is shared, and whether their information in particular has been shared with a third party.

RECOMMENDATION

Companies should prepare transparency reports using a standardized reporting template. The

[70] Parsons, Christopher. (2016). "Release: DIY Transparency Reporting Tool," *The Citizen Lab*, retrieved December 7, 2017, `https://citizenlab.org/2016/06/release-diy-transparency-report-tool/`.

[71] Parsons, Christopher. (2016). "Transparency in Surveillance: Role of various intermediaries in facilitating state surveillance transparency," Centre for Law and Democracy, retrieved December 7, 2017, `http://responsible-tech.org/wp-content/uploads/2016/06/Parsons.pdf`; Parsons, Christopher. (2015). "The Governance of Telecommunications Surveillance: How Opaque and Unaccountable Practices and Policies Threaten Canadians,"*Citizen Lab - The Telecom Transparency Project*, retrieved December 7, 2017, `https://citizenlab.ca/2015/05/governance-of-telecommunications-surveillance/`.

DIY Transparency Reporting Tool developed by the Citizen Lab,[72] offers an example of such a template based on the model issued by Industry Canada.[73]

8.1.4 DATA DEFINITIONS

An area of high variation within industries pertained to common language terms that have assumed unique meanings within each organization. "Subscriber data" or "subscriber information," as an example, are terms of art used by telecommunications companies and loosely by dating services, and can constitute radically different data items. In some cases a request using the term will elicit email addresses and customer name and home address. In other cases, the same request will also elicit a social insurance number or credit card number, and in others the IP address used to sign up to the service. Common industry understandings of what commonly-relied-upon terms mean would help consumers build confidence that they understand the kinds of data linked to these terms, and help ensure that companies have relatively coherent lawful access handbooks so that authorities, as well, would know what they are likely to obtain when they make requests. These efforts would, in effect, improve trust with customers while potentially focusing requests from law enforcement on the specific types of terms they require for their lawful investigations.

RECOMMENDATION

Organizations should collaborate within their respective industries to establish definitions for personal data terms, such as subscriber data, metadata, content of communications, etc. The goal should be to help consumers better understand how these common-language terms are used across an industry, as well as help to narrow and focus government agency requests for organization-held personal data.

8.1.5 DAR PROCESS EFFICIENCY

Each industry we examined had several examples of barriers to providing our participants with access to their personal information. Telecommunications service providers routinely asked for payment for sample or comprehensive records, such as calling records, cell location, or SMS metadata. Online dating services, while contesting that PIPEDA even applied to them as non-Canadian entities storing personal information outside Canada, had rigorous identity verification procedures that sometimes requested the participant to send notarized copies of identity documents via postal mail. Finally, two fitness tracking providers, Fitbit and Basis, required

[72] Parsons, Christopher. (2016). "Release: DIY Transparency Reporting Tool," *The Citizen Lab*, retrieved December 7, 20176, `https://citizenlab.org/2016/06/release-diy-transparency-report-tool/`.

[73] Industry Canada. (2015). "Transparency Reporting Guidelines," *Government of Canada*, retrieved December 7, 2017, `http://www.ic.gc.ca/eic/site/smt-gst.nsf/eng/sf11057.html`.

participants to use third party software to transmit data in a confidential manner and required technical skills and detailed coordination with the companies.

RECOMMENDATION

Companies should review their access processes, and assess where improvements could be made to reduce cost, reduce or make more user-friendly their identity verification steps, and streamline security procedures. Publicizing their access processes could also help requesters better prepare their initial DARs to match company requirements. Investing in software solutions to assist in fulfilling DARs may also help industries coalesce around standard DAR fulfilment workflows.

8.2 DURING REQUEST

8.2.1 CLARIFY COMMUNICATIONS CHANNELS

When companies receive DAR letters they must understand what questions are being posed by the subscriber and how best to respond. While requests made using the AMI web application are intended to be specific and self-apparent to a recipient, some organizations may be uncertain of the full implications of the question. By developing accessible and customer-focused approaches to resolving these uncertainties the customer will ideally receive a productive response more quickly than if the uncertainties are left to percolate. To resolve questions a party that has received a DAR can contact the requester to, first, ensure that they develop a communications channel that is amenable to the customer. Sometimes this will involve using email, other times telephone calls, and other times letter mail. By using a customer's preferred means of communication a organization can ensure that their customers feel empowered and in control of the communication, as opposed to being forced to communicate in a way that is distracting or upsetting.

RECOMMENDATION

Companies should not assume that they know which communications method their customers would prefer to use when discussing a DAR letter. They should first ask the customer what their preferred method is and only then pose questions to clarify the requester's inquiries.

8.2.2 RETENTION SCHEDULES

Sometimes companies do not have data that one of its subscribers is requesting. In some situations, however, organizations do not specifically assert that they lack the requested information and either instruct a subscriber to just download their data (to, presumably, discover that

the data requested is not in the downloaded file(s)) or fail to respond to the question in a substantive way. This response can be confusing because a subscriber may not know whether the information is not being provided or does not exist.

RECOMMENDATION

Organizations should either make data retention guidelines publicly available, or provide them to any subscriber asking about a company's data collection and retention policy. Where a subscriber is requesting copies of data that are not collected by a company the company should specifically affirm that it does not collect that type of data.

8.2.3 DATA INVENTORIES

PIPEDA establishes a right to compel companies to disclose personal information they have collected about their subscribers. Sometimes this data is provided directly by subscribers themselves, other times it is collected when they use a service, and other times it is gathered ambiently from the devices the organization's software is installed on. Some companies' products and services are very complicated, which can lead to personal information being scattered throughout an organization and sometimes making it challenging to collect and provide to a subscriber when they request copies of their personal information.

PIPEDA permits for some negotiation where such complications arise to help answer a subscriber's questions while minimizing costs to the organization in question. Though it may be expensive or time consuming to provide all copies of all data held about a given subscriber a company should, at the bare minimum, freely provide a few copies of each kind of data that is held to satisfy concerns about what data is collected by the organization as a result of a subscriber using the service. This should not be interpreted as the same thing as providing full access, and companies should, upon further request from their customers, provide complete access to all records containing personal information at no or minimal cost.

RECOMMENDATION

Companies should publish data inventories describing all the kinds of personal information that they collect, and freely provide copies of a small set of representative examples of records for each kind of personal information to subscribers upon request. Further, they should communicate with customers and agree upon a reasonable cost for providing information in excess of the handful of records that can be provided freely. Such costs should not be designed to, or have the effect of, disincentivizing customers from obtaining copies of their personal information.

8.2.4 CAPACITY TO UNDERSTAND DATA

Many requesters may not have the expertise to appreciate the significance of the data provided to them, leaving them unable to appreciate the meaning and context of the provided information.

RECOMMENDATION

Companies should provide access to personal data in a usable format. Data that includes rows and columns should be provided as a CSV spreadsheet or other open format. Companies should not provide any textual or numeric data in an image format, such as a screenshot or image-based PDF. Text and numbers should be easily searchable, selectable, and easy to copy.

RECOMMENDATION

People who issue DARs may benefit from tool support that can help them to better make sense of the data they get back. Researchers and programmers should develop public tools that facilitate the analysis of data retrieved through DARs. These tools should not collect any personal information, and instead process any data locally on a user's device. Companies themselves could offer similar tools within their customer portals.

8.2.5 SECURE TRANSMISSION

Several companies sent data to our participants over email without using any sort of additional security. Email by itself does not offer strong security protections. Apple, OkCupid, and Tinder all sent data over email in the form of an encrypted zip file, and sent the password for that zip file in a separate email. Basis proposed a high level of security to our participant through PGP mail encryption. However, PGP is challenging for many people to download, install, and operate. Basis' requirement that its users use PGP served as a barrier to our participant, who was unable to receive their data.

Several fitness tracking companies directed our participants to data download tools. Such tools offer a convenient and relatively secure way to access the personal data held by a company. However, in our sample, we found these data downloads were not as extensive as those provided by companies that sent data directly to participants. In particular, data export tools did not include the breadth of metadata that was provided by companies which directly sent data to individuals.

RECOMMENDATION

Companies should offer strong security protections when transmitting personal data in response to a DAR. The requester should not have to undertake significant additional effort as a result of

these security mechanisms. Companies should offer data download tools that are accessible through their existing online user portals. In their DAR responses, companies should respond to any questions asked of them and direct requesters to their data download tool.

These data export tools should provide complete access to all personal data retained about the user, including data such as access logs, location history, analytics records, account status history, and customer service interactions. The data download should be conducted over an encrypted channel. Data should be provided in a usable format. The downloaded data should be accompanied by a data dictionary that describes the meaning and context of all provided data.

8.3 AFTER REQUEST

8.3.1 REQUESTER FEEDBACK

Large and small organizations alike sometimes struggle in their responses to customers' DAR letters. They might not have received a letter making similar kinds of inquiries before, staff might be unable to answer questions, comprehensively, and responses might be influenced by, but fail to make clear, an interpretation of PIPEDA's requirements. These struggles are normal but can be made fruitful when organizations actively re-evaluate the levels of satisfaction that DAR requesters have to their responses and, subsequently, make changes to better respond to the inquiries.

RECOMMENDATION

Organizations should engage in follow-up surveys to determine whether subscribers are satisfied with the disclosures they have received.

RECOMMENDATION

Organizations should modify their data disclosure practices to alleviate issues or concerns that subscribers note in surveys that evaluate whether customers are satisfied with the disclosures they have received.

8.3.2 STAKEHOLDER ENGAGEMENT

Non-corporate stakeholders, such as the Citizen Lab, have developed online tools to help customers create DAR letters. These stakeholders modify request letter language from time to time based on new research or feedback from industry groups. By engaging with these kinds of stakeholders, organizations can improve on the DAR process and potentially reduce the challenge in responding to customers' requests.

RECOMMENDATION

Either individual organizations or industry groups should communicate with non-corporate stakeholders to help streamline the request process, or to help establish requesters' expectations. This might involve developing Application Programming Interfaces (APIs) to expedite the issuance and response to DAR letters or working to modify language used by web applications to more accurately reflect the data that might be held by organizations.

9 FUTURE WORK

To date, the Citizen Lab has used DAR letters to better understand the data collection, retention, and processing activities which are undertaken by organizations in three separate industries. What we have learned about each organization has varied significantly, and revealed both differences in the quality of information presented by individual organizations as well as variances across industries. PIPEDA has been in force for over a decade, but organizations located in Canada and in foreign jurisdictions alike often provide unsatisfactory responses or, in the case of foreign organizations, take the position that the law does not apply to them.

Future work may use interviews to deepen researchers' understanding of how industry members respond to these types of DAR letters. Semi-structured interview questions could probe how companies currently respond to the letters, internal views of these kinds of requests, as well as challenges companies face in responding to requests. Questions could also focus on whether the issuance of DARs have prompted changes in corporate culture concerning data handling. In aggregate, the questions would shed light on whether DAR letters are effective in revealing corporate activity while also helping external stakeholders develop tools to relieve unnecessary pressures on companies which receive these requests.

Interviews could also be held with privacy and data protection commissioners and professionals to understand how privacy-focused stakeholders perceive DAR letters being generated using the AMI web application. In the case of Commissioners or Commissioners' staff, a semistructured interview guide could be designed to probe whether the structuring of questions is useful for resolving disputes between organizations and their subscribers. Questions could also focus on whether Commissioners think the widespread issuance of DARs has influenced corporate culture with regard to data handling practices (e.g. are companies more mindful of information they retain, with whom it is shared, etc) or government understandings of PIPEDA's application. This latter question in particular would be used to triangulate whether the Commissioner believes DAR requests are a useful way of modifying corporate behaviour or whether alternate processes are needed instead. Interviews with privacy-focused stakeholders, such as NGOs or academics, might ask them to evaluate the impact of DAR letters or whether their issuance is helpful for encouraging corporate transparency.

To date, the providers included in the AMI web application have been digital communications and technology companies. These kinds of organization were selected based on the sensitivity of the personal information that they access. However, it remains to be seen whether other types of companies, such as insurance companies, banks, airlines, or other companies that handle significant amounts of personal information, would necessarily provide a more or less satisfactory set of responses to DAR inquiries. Research examining other sectors would help to show whether the challenges subscribers have in obtaining detailed answers to their ques-

tions is restricted to the few industry types that we have examined thus far or whether, instead, there is a cross-economy failure by organizations to satisfactorily respond to customers' DAR letters.

Our analysis in this report has been focused on how companies respond to DARs that we developed. The core template for the DARs was developed in 2014. Since that time, we have collected data on how many different companies respond to these letters. Future research could examine the limitations we identified in how companies respond and assess how the DAR letter itself could be modified to mitigate or minimize some of these limitations. For example, research could examine whether or not it would be helpful to provide an example of a "model response" to a DAR as an attachment to future DARs, which would provide guidance to companies regarding how we would imagine the DAR questions could be answered. Research could examine whether or not providing more detail, as a link or attachment, about what exactly is meant by the various data types we request would lead to more fulsome responses, and could include providing definitions, examples, and model responses for each data type. Such studies could potentially help requesters better understand the data they are requesting as well as companies to respond more completely and in greater detail.

10 CONCLUSION

We have found that DARs can help clarify the collection, retention, and disclosure of personal information. However, significant barriers exist to obtaining full responses to DARs. Barriers are present in each industry we examined, but differ between them: cost is the primary barrier for telecommunications, identity verification for online dating, and data transfer security for fitness tracking. Non-Canadian fitness tracking and online dating companies sometimes failed to respond to DARs at all, or questioned the applicability of Canadian privacy law to their operations. Even if those barriers are overcome, the substance of DAR responses still pose challenges to those seeking to better understand how companies treat their personal information.

Our prior work into fitness tracking industry demonstrated that DAR responses can omit references to entire categories of data that we observed being transmitted to companies, which leads us to hypothesize that requesters cannot be certain that the data they receive in response to a DAR is necessarily a complete record of what is retained by the company. In order to get a more complete understanding of data retained by companies, technical measurement of data transmissions coupled with privacy policy analysis can provide important supplementary data.

DARs did shed some light on variation between industries regarding data retention periods. While no company explicitly provided data retention schedules to help requesters understand what data could be available for them to obtain, telecommunications companies were more likely to make claims about retention periods for specific types of information. For fitness tracking and dating industries, we found examples of seemingly indefinite retention insofar as companies in both categories provided requesters with logs of IP addresses, granular heart rate records, and timestamped precise geolocation records that all dated back to the time when our participant requesters created their accounts.

While DARs have limitations, companies can take steps to provide more complete access to data and clearer responses to requesters' questions. There is correspondingly more that we, as the principal authors of the request letters used in this study and by thousands utilizing the AMI web application, can do to improve the letters to guide companies to provide access more readily. By addressing these recommendations and areas of future work, it is our hope that the DAR process can be improved for companies and for requesters exercising their right to learn about their personal information.

Ultimately, DARs provide a valuable new method for understanding the kinds of information which are collected, retained, processed, and handled by private companies. And this report constitutes the first time that a research institution has evaluated how companies respond to these access rights, and drawn lessons both within specific industry groupings and across industries. Given the amounts of digital information that individuals produce on a daily basis it is imperative that they can gain access to such information upon request, especially when

companies rarely public information concerning their data collection, retention, handling, or disclosure practices. Our report showcases that DARs can provide insight into private practice but, as of today, the DAR processes themselves are immature and in need of either private leadership to advance individuals' access to their personal information, or sector-wide responses to compel changes in how private organizations provide access to the information of which they are stewards.

A AMI PROJECT REQUEST LETTERS

SAMPLE TELECOMMUNICATIONS SERVICE PROVIDER LETTER

March 28th, 2017
Chief Privacy Officer, Rogers Group of Companies
333 Bloor Street East
Toronto
M4W 1G9

Dear Privacy Officer:

I am a user of your telecommunications service, and am interested in both learning more about your data management practices and about the kinds of personal information that you maintain and retain about me. So this is a request to access my personal data under' Principle 4.9 of Schedule 1 and section 8 Canada's federal privacy legislation, the Personal Information Protection and Electronic Documents Act (PIPEDA).

I have the following questions about the collection, use, and disclosure of my personal data:

I am requesting a copy of all records which contain my personal information from your organization. The following is a non-exclusive listing of all information that Rogers may hold about me, including the following:

- **Call logs** (e.g. numbers dialed, times and dates of calls, call durations, routing information, and any geolocational or cellular tower information associated with the calls)
- **Mobile app data** Information collected about me, or persons/devices associated with my account, using one of your company's mobile device applications
- **Geolocation data** collected about me, my devices, and/or associated with my account (e.g. GPS information, cell tower information)
- **IP address logs** associated with me, my devices, and/or my account (e.g. IP addresses assigned to my devices/router, IP addresses or domain names of sites I visit and the times, dates, and port numbers)
- **Disclosures to third parties** Any information about disclosures of my personal information, or information about my account or devices, to other parties, including law enforcement and other state agencies
- **Text & multimedia messages** (sent and received, including date, time, and recipient information)
- **Subscriber information** that you store about me, my devices, and/or my account

- **Other** Any additional kinds of information that you have collected, retained, or derived from the telecommunications services or devices that I, or someone associated with my account, have transmitted or received using your company's services

If your organization has other information in addition to these items, I formally request access to that as well. If your service includes a data export tool, please direct me to it, and ensure that in your response to this letter, you provide all information associated with me that is not included in the output of this tool. Please ensure that you include all information that is directly associated with my name, phone number, e-mail, or account number, as well as any other account identifiers that your company may associate with my personal information.

You are obligated to provide copies at a free or minimal cost within thirty (30) days in receipt of this message. If you choose to deny this request, you must provide a valid reason for doing so under Canada's PIPEDA. Ignoring a written request is the same as refusing access. See the guide from the Office of the Privacy Commissioner at: `http://www.priv.gc.ca/information/guide_e.asp#014`. The Commissioner is an independent oversight body that handles privacy complaints from the public.

Please let me know if your organization requires additional information from me before proceeding with my request.

Here is information that may help you identify my records:

- First Name: Example
- Last Name: Example
- Address 1: 123 Example Street
- City: Exampleville
- Province: New Brunswick
- Postal Code: X1X 2X2
- Email Address: Example@Example.com
- Telephone Number: (xxx)-xxx-xxxx
- Account Number: 123EXAMPLE567

Sincerely,
Example Example

SAMPLE ONLINE DATING COMPANY LETTER

March 28th, 2017
Privacy Officer, OkCupid.com
8300 Douglas Avenue, Suite 800
Dallas
75225

Dear Privacy Officer:

I am a user of your dating application, and am interested in both learning more about your data management practices and about the kinds of personal information that you maintain and retain about me. So this is a request to access my personal data under Principle 4.9 of Schedule 1 and section 8 of Canada's federal privacy legislation, the Personal Information Protection and Electronic Documents Act (PIPEDA).

I am, first of all, requesting more information about how data is collected and exchanged by you with other companies or organizations. Can you clarify whether my data, either in an individualized data set or part of an aggregate data set, has been provided to other parties? And if it has been provided (either voluntarily, as part of a commercial transaction, or on other grounds) please identify to which parties it has been provided.

Second, I wanted to understand a bit more about how my data could be disclosed to government authorities. What are your specific policies, practices, or processes for handling requests from authorities from international jurisdictions, such as from Canadian policing organizations? How would you respond if my information was requested as evidence in a Canadian civil or criminal proceeding?

Third, is the data transmitted between the application of yours that I have installed and your servers secured against potential eavesdroppers?

Finally, I am requesting a copy of all records which contain my personal information from your organization. The following is a non-exclusive listing of all information that OkCupid may hold about me, including the following:

- **Geolocation data** collected about me, my devices, and/or my account
- **Any additional kinds of information** that you have collected, retained, or derived from the mobile or website services you provide, including by not limited to: data or records collected using my camera or from my camera roll; social networking information; data collected or retained derived from my microphone; or communications between myself and other users; contact book information;

- **Lifestyle information** that you may have about me, such as drinking habits or sexual preference information.
- **Personally identifying information** that is unique to me, my devices, and/or my account, such as name, email addresses, phone numbers, responses to relationship questions, or device identifiers;
- **Mobile app data** Information collected about me, or persons/devices associated with my account, using one of your company's mobile device applications
- **IP address logs** associated with me, my devices, and/or my account (e.g. IP addresses assigned to my devices/router, IP addresses or domain names of sites I visit and the times, dates, and port numbers)
- **Subscriber information** that you store about me, my devices, and/or my account

If your organization has other information in addition to these items, I formally request access to that as well. If your service includes a data export tool, please direct me to it, and ensure that in your response to this letter, you provide all information associated with me that is not included in the output of this tool. Please ensure that you include all information that is directly associated with my name, phone number, e-mail, or account number, as well as any other account identifiers that your company may associate with my personal information.

You are obligated to provide copies at a free or minimal cost within thirty (30) days in receipt of this message. If you choose to deny this request, you must provide a valid reason for doing so under Canada's PIPEDA. Ignoring a written request is the same as refusing access. See the guide from the Office of the Privacy Commissioner at: `http://www.priv.gc.ca/information/guide_e.asp#014`. The Commissioner is an independent oversight body that handles privacy complaints from the public.

Please let me know if your organization requires additional information from me before proceeding with my request.

Here is information that may help you identify my records:

- First Name: Example
- Last Name: Example
- Email Address: Example@Example.com
- Telephone Number: (xxx)-xxx-xxxx
- Username: Example

Sincerely,
Example Example

SAMPLE FITNESS TRACKER COMPANY LETTER

March 28th, 2017

Dear Privacy Officer:

I am a user of your fitness tracking device, and am interested in both learning more about your data management practices and about the kinds of personal information that you maintain and retain about me. So this is a request to access my personal data under Principle 4.9 of Schedule 1 and section 8 of Canada's federal privacy legislation, the Personal Information Protection and Electronic Documents Act (PIPEDA).

I am, first of all, requesting more information about how data is collected and exchanged by you with other companies or organizations. Can you clarify whether my data, either in an individualized data set or as part of an aggregate data set, has been provided to insurance agencies? And if it has been provided (either voluntarily, as part of a commercial transaction, or on other grounds) please identify to which insurance agencies it has been provided.

Second of all, I wanted to clarify what jurisdiction any concerns, complaints, or conflicts are resolved in. I live in Canada; am I bound to engage with your company in a non-Canadian arbitration or legal environment? I am not planning on engaging in such a conflict but wanted to better understand my rights.

Third, I wanted to understand a bit more about how my data could be disclosed to government authorities. What are your policies, practices, or processes for handling requests from authorities from international jurisdictions, such as from Canadian policing organizations? How would you respond if my information was requested as evidence in a Canadian court case or criminal proceeding?

Fourth, is the personal data transmitted between my mobile phone and your web servers secured against potential eavesdroppers? What about between my fitness band and my phone?

Fifth, can you describe in more detail what practices you've implemented to ensure Bluetooth data transmissions are privacy-protective?

Finally, I am requesting a copy of all records which contain my personal information from your organization. The following is a non-exclusive listing of all information that Apple may hold about me, including the following:

- **Geolocation data** collected about me, my devices, and/or my account
- **Any additional kinds of information** that you have collected, retained, or derived from the mobile or website services your provide, or with the fitness-related device your company produces that I use

- **Health and fitness data** including all records of my step activity, heart rate, sleep patterns, food intake.
- **Mobile app data** Information collected about me, or persons/devices associated with my account, using one of your company's mobile device applications
- **IP address logs** associated with me, my devices, and/or my account (e.g. IP addresses assigned to my devices/router, IP addresses or domain names of sites I visit and the times, dates, and port numbers)
- **Disclosures to third parties** Any information about disclosures of my personal information, or information about my account or devices, to other parties, including law enforcement and other state agencies
- **Subscriber information** that you store about me, my devices, and/or my account

If your organization has other information in addition to these items, I formally request access to that as well. If your service includes a data export tool, please direct me to it, and ensure that in your response to this letter, you provide all information associated with me that is not included in the output of this tool. Please ensure that you include all information that is directly associated with my name, phone number, e-mail, or account number, as well as any other account identifiers that your company may associate with my personal information.

You are obligated to provide copies at a free or minimal cost within thirty (30) days in receipt of this message. If you choose to deny this request, you must provide a valid reason for doing so under Canada's PIPEDA. Ignoring a written request is the same as refusing access. See the guide from the Office of the Privacy Commissioner at: `http://www.priv.gc.ca/information/guide_e.asp#014`. The Commissioner is an independent oversight body that handles privacy complaints from the public.

Please let me know if your organization requires additional information from me before proceeding with my request.

Here is information that may help you identify my records:

- First Name: Example
- Last Name: Example
- Email Address: Example@Example.com

Sincerely,
Example Example

B AMI USER SURVEY QUESTIONS

1. What company did you request your data from?
    - _____
2. How have you learned about the privacy practices of the company above?
    - Privacy Policy
    - Social Media
    - Customer Service
    - News reports
    - People I trust
    - Other
3. How familiar were you with your right of access to personal information before using Access My Info?
    - Scale of 1 to 5
4. What were you interested in learning about by requesting your data?
    - _____
5. On what date did you send in your request?
    - _____
6. Did the company you requested data from respond in any form?
    - Yes
    - No
7. What communications channel did they use to respond?
    - E-mail
    - E-mail with attachment
    - E-mail with link to download page
    - Postal mail
    - Registered mail (signature required)
8. Did they respond within 30 days?

- Yes
- Yes, but to ask for an extension
- No

9. Did they provide you with your data?
    - Yes
    - Yes, but only after some back and forth
    - No
10. Did they answer the questions asked?
    - Yes
    - Yes, but only after some back and forth
    - No
11. Did they charge a fee?
    - Yes
    - No
12. How much did they charge for access?
    - _____
13. Did you pay it?
    - Yes
    - I negotiated a lower fee for less data, and paid that
    - No
14. Did they ask for extra identity verification?
    - Yes
    - No
15. What did they want you to provide to verify your identity?
    - _____
16. On a scale of 1 to 5, how much did the response you received help you understand how your data is used?

- Scale of 1 to 5

17. If any, what data do you feel was missing from the response?
    - _____
18. Did you find Access My Info easy to use?
    - Yes
    - No
19. What challenges did you face using Access My Info?
    - _____
20. Would you recommend Access My Info to a friend?
    - Yes
    - No
21. What type of organization would you like to see Access My Info support next?
    - _____
22. Do you have any other comments about Access My Info?
    - _____