

New Data for Access to Justice:  
An Assessment of the value of Reddit data

by

Katherine M. Gower

B.Sc. (Honours), University of Victoria, 1994  
M.A. (Journalism), University of Western Ontario, 1997  
J.D., University of Victoria, 2008

A Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of

MASTER OF LAWS

in the Faculty of Law

© Katherine M. Gower, 2024  
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

We acknowledge and respect the Lək̓ʷəŋən (Songhees and Esquimalt) Peoples on whose territory the university stands, and the Lək̓ʷəŋən and W̱SANEĆ Peoples whose historical relationships with the land continue to this day.

New Data for Access to Justice:  
An Assessment of the value of Reddit data

by

Katherine M. Gower

B.Sc. (Honours), University of Victoria, 1994  
M.A. (Journalism), University of Western Ontario, 1997  
J.D., University of Victoria, 2008

Supervisory Committee:

**Jeremy Webber, FRSC, Co-Supervisor**  
Faculty of Law

**Robert G.W. Lapper, K.C., Co-Supervisor**  
School of Public Administration

## ABSTRACT

Canada has an “access to justice” problem. Research shows that most Canadians will experience a legal problem in a three-year period, and less than half of them will resolve their problem in that time. The justice system wants to improve this situation, but it needs evidence and data regarding what problems people have, and the impact of any changes made, in order to do so. In this thesis, I assess the empirical data available, introduce a new form of data, and provide an assessment of both. The new data comes from the social media platform Reddit, and it was uncovered and collected by the interdisciplinary team at a new incubator lab - the Justice Data and Design Lab (the JDD Lab), which I directed.

The JDD Lab used machine learning and artificial intelligence to locate, collect and analyze Reddit data from a public “subreddit” regarding legal advice for Canadians. The data are first-person posts which represent a unique opportunity to hear descriptions from real people who find themselves facing what they either know or suspect is a legal problem and reach out online for help. Analysis shows Reddit data clusters into some of the same topic areas used in Everyday Legal Needs surveys, but also generates new clusters which represent real-time concerns that the public are experiencing.

This thesis juxtaposes Everyday Legal Needs data and Reddit data to assess their strengths and weaknesses. It concludes that if the justice system is serious about taking a person-centred approach to improving access to justice, then Reddit data is an essential complement to Everyday Legal Needs survey work.

Table of Contents

**SUPERVISORY COMMITTEE:..... II**  
**ABSTRACT.....III**  
**TABLE OF CONTENTS ..... IV**  
**LIST OF FIGURES ..... V**  
**LIST OF TABLES ..... VI**  
**ACKNOWLEDGEMENTS ..... VII**  
**CHAPTER 1: CONTEXT - THE ACCESS TO JUSTICE PROBLEM ..... 1**  
    INTRODUCTION ..... 1  
    ACCESS TO JUSTICE ..... 1  
    THE DATA DEFICIT ..... 8  
    MY ROLE ..... 13  
**CHAPTER 2: EVERYDAY LEGAL NEEDS SURVEYS AND DATA ..... 16**  
**CHAPTER 3: THE JUSTICE DATA AND DESIGN LAB (THE JDD LAB) AND REDDIT DATA ..... 35**  
    COLLECTION ..... 36  
    REDDIT ..... 39  
    REDDIT DATA COLLECTION AND AGGREGATION ..... 40  
    ANALYSIS ..... 43  
    RESULTS ..... 46  
**CHAPTER 4: ASSESSMENT ..... 49**  
    EVERYDAY LEGAL NEEDS DATA ..... 49  
    REDDIT DATA ..... 56  
    ASSESSMENT ..... 57  
**CHAPTER 5: CONCLUSION AND FUTURE WORK ..... 69**  
    CONCLUSIONS ..... 69  
    FUTURE WORK ..... 71  
**APPENDIX 1 ..... 74**

## List of Figures

Figure 1: Google Trends results for “legal advice” (Spring 2021).....	38
Figure 2: SEMRUSH “Keyword Magic Tool” results for “legal advice” (Spring 2021)....	38
Figure 3: An Example Post from “r/legaladvicecanada” (edited for confidentiality).....	44
Figure 4: Cluster 4 - Results of LDA and Inter-topic Distance Mapping on Reddit Data.....	45
Figure 5: Cluster 5 - Results of LDA and Inter-topic Distance Mapping on Reddit Data.....	46

## List of Tables

Table 1: Most Common and Most Serious Problems in Canada and BC.....	30
Table 2: Top Five words in Clusters and Possible Topics .....	48

## Acknowledgements

I extend my gratitude to the data science scholars - Juan Roesel, Shounak Mondal and Echo Zhiang - who developed the programing that collects, stores and analyzes Reddit data. I also thank Kaitlyn Cummings, PhD Candidate in Law, for her domain expertise in law.

My thanks to Jeremy Webber and Robert Lapper, my co-supervisors, for their joyful support of my work and this thesis. I am grateful for their commitment to guiding me through this process.

Thank you to Michelle Lawrence for her enthusiasm and support of the Justice Data and Design Lab and to Heather Chestnut whose much-admired and appreciated skills with funding and budgeting made the work possible.

Thank you to John Jacobson, who offered his help when I needed it to make the Lab a reality.

This work was supported by funding from the Law Foundation of British Columbia, Canada, and matching funds from Mitacs.

## CHAPTER 1: CONTEXT - THE ACCESS TO JUSTICE PROBLEM

### Introduction

Canada has an “access to justice” problem. Research shows that most Canadians will experience a legal problem in a three-year period, and less than half of them will resolve their problem in that time. National committees working on the problem say that the justice system is “too complex, too slow and too expensive”<sup>1</sup>, and that it no longer provides proportional and just outcomes for those who seek to use it. Considerable work is underway to improve this situation. A key part of the solution involves gathering and using evidence and data regarding what problems people have, and the impact of any changes made. In this thesis, I assess the empirical data available, introduce a new form of data, and provide an assessment of both. The new data comes from the social media platform Reddit, and it was uncovered and collected by the interdisciplinary team at a new incubator lab - the Justice Data and Design Lab – which I directed. This introductory chapter sets the background and the context for the work undertaken in this thesis.<sup>2</sup>

### Access to Justice

Access to justice issues are not unique to Canada. While this thesis focusses on access to justice work in civil justice in Canada, and more specifically in the province of British Columbia, work in this field is underway internationally. In the 1970s, a four-year, access to justice project

---

<sup>1</sup> Action Committee on Access to Justice in Civil and Family Matters, “Access to Civil & Family Justice: A Roadmap for Change” (2013) Ottawa, Canada (“Canada’s Roadmap”), online (pdf): *Canadian Forum on Civil Justice* < [https://www.cfcj-fcjc.org/sites/default/files/docs/2013/AC\\_Report\\_English\\_Final.pdf](https://www.cfcj-fcjc.org/sites/default/files/docs/2013/AC_Report_English_Final.pdf) > [<https://perma.cc/M5QR-ZWLX>].

<sup>2</sup> Access to Justice work and scholarship is an evolving field. For this thesis, I cap my review of reference materials at the year of the reported JDD Lab research, i.e. 2021. I make one exception and include a 2022 report on the 2021 Canadian Legal Problems Survey.

prepared a four-volume report which included a review of access to justice problems worldwide, and studies of promising reform efforts around the world.<sup>3</sup> In 1992, the United Nations began work towards a blueprint for sustainable worldwide development. In 2015, United Nations member states adopted an international agenda which included 17 Sustainable Development Goals (SDGs). Goal 16 is to “[p]romote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels”<sup>4</sup>.

In England, successive waves of access to justice work have washed over the isles. Early on, concerns about the high costs of accessing the law led to the creation of financial aid for people who needed to use the justice system.<sup>5</sup> “Legal aid”, as it is now generally known, has been challenged, discussed and (sometimes) improved ever since. In the 1990s, the Woolf Report reviewed the rules of court, with the goal of improving access to justice by reducing the costs and complexity of litigation.<sup>6</sup> National surveys were undertaken by the National Consumer Council and by researcher Hazel Genn (with funding from the Nuffield Foundation) which provided information on people’s satisfaction with the justice system, and on the kinds of problems people were experiencing outside the formal justice system.<sup>7</sup>

---

<sup>3</sup> Mauro Cappelletti and Bryant Garth, “Access to Justice: The Newest Wave in the Worldwide Movement to Make Rights Effective” (1978) 27 *Buff L Rev* 181 (“Cappelletti and Garth (1978)”). B. Makowiecky Salles and P. Márcio Cruzl, “The Florence Access-To-Justice Project: Descriptive Aspects” (2022) 22 *Revista de Derecho* 178.

<sup>4</sup> United Nations, Department of Economic and Social Affairs, Sustainable Development - UN Goal 16, online: *United Nations Sustainable Development Goals* <<https://sdgs.un.org/goals/goal16>> [<https://perma.cc/56X3-ZAKZ>].

<sup>5</sup> L.C.B. Gower, “The Future of the English Legal Profession” (1946) 9:3 *Mod L Rev* 211; L.C.B. Gower, “Interim Report of Committee on Supreme Court Practice and Procedure (Cme 7764)” (1949) 12:4 *Mod L Rev* 483; L.C.B. Gower, “The Cost of Litigation: Reflections on the Evershed Report” (1954) 17:1 *Mod L Rev* 1.

<sup>6</sup> A.A.S. Zuckerman, “Lord Woolf’s Access to Justice: Plus ça change ...” (1996) 59:6 *Mod L Rev* 773.

<sup>7</sup> National Consumer Council, *Seeking Civil Justice: A Survey of People’s Needs and Experiences* (London: National Consumer Council, 1995); Hazel Genn, *Paths to Justice: What People Do and Think About Going to Law* (London: Hart Publishing, 1999) (“Genn 1999”).

In the 2000s, national surveys continued to provide information on people’s legal needs.<sup>8</sup> Significant cuts to legal aid were legislated in 2013, and a flurry of reports catalogued the impact on access to justice. The Law Society, Amnesty International UK and the Bach Commission respectively reported that the cuts meant that legal aid was not available for many who needed it, that they were impacting human rights, and that there were fundamental things “wrong” with the justice system which needed to be fixed.<sup>9</sup> At the same time, the Ministry of Justice launched a £1 billion program to modernize and reform the country’s courts and tribunals.<sup>10</sup> Her Majesty’s Courts and Tribunal Service, the Administrative Justice Council and the Legal Education Foundation collaborated to develop and report on a data strategy which could support the empirical testing, reviewing and improving of the targeted systems.<sup>11</sup> Yet, the challenge of how to improve access to justice persists.

---

<sup>8</sup> Pascoe Pleasence and Nigel Balmer, *How People Resolve Legal Problems* (Cambridge: Legal Services Board, 2014); The Legal Services Society and The Law Society, “Legal Needs of Individuals in England and Wales – Summary Report 2019/20” (2020), online (blog with links): *The Law Society* <<https://www.lawsociety.org.uk/topics/research/legal-needs-of-individuals-in-england-and-wales-report>> [<https://perma.cc/7DB3-WPJP>]

<sup>9</sup> The Law Society, “Access Denied? LASPO Four Years On: A Law Society Review” (2017), online (blog with download links): *The Law Society* <<https://www.lawsociety.org.uk/topics/research/laspo-4-years-on>> [<https://perma.cc/JE4B-ZHVH>]; Amnesty International UK, “Cuts that Hurt: The Impact of Legal Aid Cuts in England on Access to Justice” (2016), online (pdf): *Amnesty International UK* [https://amnesty.org.uk/files/aiuk\\_legal\\_aid\\_report.pdf](https://amnesty.org.uk/files/aiuk_legal_aid_report.pdf) [<https://perma.cc/HZW8-7L8K>]; England, The Bach Commission on Access to Justice, *The Crisis in the Justice System in England & Wales – Interim Report* (London: Fabian Society, 2016); England, The Bach Commission, *The Right to Justice: The Final Report of the Bach Commission* (London: Fabian Society, 2017).

<sup>10</sup> England and Wales, Lord Chancellor, the Lord Chief Justice and the Senior President of Tribunals, “Transforming Our Justice System” (Ministry of Justice, 2016), online (pdf): *Government of the UK* <[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/553261/joint-vision-statement.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/553261/joint-vision-statement.pdf)> [<https://perma.cc/KV8F-2YWS>].

<sup>11</sup> Senior President of Tribunals (2018) “The Modernisation of Tribunals 2018: A Report by the Senior President of Tribunals” at 9, online (pdf): *Courts and Tribunals Judiciary* <[https://www.judiciary.uk/wp-content/uploads/2019/01/6.5332\\_JO\\_Modernisation-of-Tribunals-2018-Report\\_v3.pdf](https://www.judiciary.uk/wp-content/uploads/2019/01/6.5332_JO_Modernisation-of-Tribunals-2018-Report_v3.pdf)> [<https://perma.cc/GV5R-8ACK>].

In Australia, work is underway at both the national and state level. In 2018, the Law Council of Australia reviewed the state of access to justice country-wide and provided a roadmap for action to improve the situation.<sup>12</sup> At the same time, the Law Foundation of New South Wales and the Victoria Law Foundation were both engaging in significant research projects. For example, the Law Foundation of New South Wales conducted a 2012 Legal Australia-Wide Survey on legal needs and prepared a 2014 Legal Assistance Service Data Digest. In 2020, it used these two datasets to provide “snapshots” of the housing-related and employment-related legal needs of different Australian populations and the services available to meet those needs.<sup>13</sup> In the adjacent state, the Victoria Law Foundation undertook a “Public Understanding of the Law Survey”, which seeks to learn more about how people understand the law and engage with it.<sup>14</sup>

In Canada, the National Action Committee on Access to Justice in Civil and Family Matters (the “National Committee”) divided the problem into four key areas and set up working groups to review them. The areas were: improving access for the public, simplifying court processes, improving family law, and prevention, triage, and referral. In 2013, the National Committee pulled together the work by all four working groups, and released its final report, “A Roadmap

---

<sup>12</sup> Law Council of Australia, “The Justice Project Final Report – Introduction and Overview” (August 2018), online (pdf): *The Law Council of Australia* <[https://lawcouncil.au/files/web-pdf/Justice%20Project/Final%20Report/Justice%20Project%20\\_%20Final%20Report%20in%20full.pdf](https://lawcouncil.au/files/web-pdf/Justice%20Project/Final%20Report/Justice%20Project%20_%20Final%20Report%20in%20full.pdf)> [<https://perma.cc/N2GR-456A>].

<sup>13</sup> Sarah Randell and Catriona Mirrlees-Black “Employment-Related Problems and Inquiries to Legal Assistance Services in NSW (Data Insights in Justice)” (2020), online: *Law and Justice Foundation of New South Wales* <<http://www.lawfoundation.net.au/ljf/app/&id=E9D3C63684ED72908525869C0004B38E>> [<https://perma.cc/Z8WK-P2PU>]; D. Bellerose and C. Mirrlees-Black “Housing-Related Problems and Enquiries to Legal Assistance Services in NSW (Data Insights In Justice)” (2020), online: *Law and Justice Foundation of New South Wales* <<http://www.lawfoundation.net.au/ljf/app/&id=9A3246E23DA309FB8525869C001133FF>> [<https://perma.cc/4QU3-QNAH>].

<sup>14</sup> Primary researchers for the Public Understanding of the Law Survey are Professor Nigel J. Balmer and Dr Hugh M. McDonald, online: *Public Understanding of the Law* <<https://puls.victorialawfoundation.org.au>> [<https://perma.cc/9QBZ-LU74>].

for Change”<sup>15</sup> (“Canada’s Roadmap”). Canada’s Roadmap called for a change in the approach to improving access to justice which included incorporating the public’s perceptions in future reforms. It contained a list of improvements needed in the formal system, including improving early resolution services, and making court processes more efficient and user-friendly.<sup>16</sup>

Canada’s Roadmap respectfully noted that while the practise of preparing reports on what was wrong with the system and what could be improved was valuable, going forward it was time to take action. It created nine Justice Development Goals (the “JDGs”) which it continues to use to track and support nation-wide action.<sup>17</sup>

Canada’s provinces and territories took up the call to action. In British Columbia, individuals and organizations in the justice sector came together to form Access to Justice BC (“A2JBC”), a collaboration committed to systemic change in the justice system. It developed an “A2JBC Approach”, supporting and amplifying actions that were user-centred, collaborative, evidence-based, and experimental. It adopted an Access to Justice Triple Aim, and created an Access to Justice Measurement Framework.<sup>18</sup> In Saskatchewan, the Centre for Research, Evaluation, and Action Towards Equal Justice (“CREATE Justice”), set three guiding principles to identify projects it would undertake: testing and reviewing design methods, implementing a collaborative, action-oriented approach and advancing justice data and metrics.<sup>19</sup> It workshopped A2JBC’s Access to Justice Measurement Framework and eventually used it to

---

<sup>15</sup> Canada’s Roadmap, *supra* note 1.

<sup>16</sup> Canada’s Roadmap, *supra* note 1 at 11 and 16.

<sup>17</sup> “Action Committee on Access to Justice in Civil and Family Matters”, online: *Justice Development Goals* <<https://www.justicedevelopmentgoals.ca>> [<https://perma.cc/TBF7-HVXX>].

<sup>18</sup> “The A2JBC Story” (undated), online: *Access to Justice BC* <<https://accesstojusticebc.ca/2020/06/the-a2jbc-story/>> [<https://perma.cc/7SKP-4JFG>].

<sup>19</sup> “CREATE Justice” (undated), online: *CREATE Justice* <https://law.usask.ca/createjustice/> [<https://perma.cc/2UY8-5RF>].

evaluate one of its projects.<sup>20</sup> These initiatives sought to create a climate of change which aligned with the National Committee’s new approach, but left the details of projects to those who undertook them.

Quebec’s *Accès au droit et à la justice*<sup>21</sup> (“ADAJ”) is a good example of the wide spectrum of projects which could improve access to justice. ADAJ parses its access to justice work into three areas: “knowledge and awareness of law as being elements of citizenship; adapting professional practices and institutional constraints in the field of justice to the actual state of social relationships; and the public and political legitimacy of contemporary legal and judicial institutions.”<sup>22</sup> To address these areas, ADAJ has over 20 “research hubs”. The research hubs include both empirical studies and innovative pilot projects undertaken in collaboration with a robust list of institutional, professional, academic and citizen partners.

With so much work underway both at the international and national level, it is perhaps not surprising that the work to improve access to justice has evolved over the years. Early work generally focussed on making the existing legal institutions equally accessible to all, no matter a person’s financial situation.<sup>23</sup> This work is still valid and necessary. Indeed, Canada’s Roadmap listed key improvements needed in the formal justice system and in its institutions. The extensive “research hubs” at Quebec’s ADAJ similarly include valuable work towards making court

---

<sup>20</sup> Guiding Principle 3, CREATE Justice, online: <https://law.usask.ca/createjustice/guiding-principle-3.php> [<https://perma.cc/G8J9-SXKJ>]; Enough Already, online: <https://law.usask.ca/createjustice/projects/ongoing-initiatives/enough-already.php> [<https://perma.cc/3PT6-4N9P>].

<sup>21</sup> In English, the *Accessing Law and Justice* research project.

<sup>22</sup> *Accès au Droit et à la Justice* (undated), online: *ADAJ* <<https://www.adaj.ca/home>> [<https://perma.cc/5FZ2-UQUX>].

<sup>23</sup> Cappelletti and Garth (1978), *supra* note 3 at 182. See also: L.C.B. Gower, *The Cost of Litigation: Reflections on the Evershed Report* (1954) 17:1 Mod L Rev 1 at 1-2.

dockets accessible, gathering justice data and statistics, considering the function of the expert witness, and advocating for “paperless justice”.<sup>24</sup> Improving the current institutional processes can reasonably be expected to improve access to justice for the people who need to use those services, and the justice system does well to continually seek to improve its offerings.

However, in the last 10 years a new perspective on improving access to justice has emerged. Canada’s Roadmap, while advocating for improvements to the formal system, also heralded this new “person-centred” approach. It called for improvements that “put the public first” and generated “user-centred” changes.<sup>25</sup> The new perspective had bubbled up in the National Committee’s working group which considered prevention, triage, and referral. That working group considered legal problems other than those which used the formal justice system. It recognized “the breadth and depth of problems in people’s everyday lives, and the need to develop a wide range of appropriate responses to these problems, rather than to funnel them to a single, high cost destination (the courts and legal representation).”<sup>26</sup> The group relied on the evidence of the “breadth and depth of problems in people’s everyday lives” which had been uncovered by surveys in England and Canada. Those surveys provided empirical evidence of the volume and type of “everyday” legal problems which people experienced and managed without ever engaging with the formal justice system. For ease of reference, I shall refer to these as “Everyday Legal Needs” surveys.

---

<sup>24</sup>Accès au Droit et à la Justice, Chantiers (undated), online: *ADAJ* <<https://www.adaaj.ca/chantiers>> [<https://perma.cc/PA8A-AV7J>].

<sup>25</sup> Canada’s Roadmap, *supra* note 1 at 2, 6-7.

<sup>26</sup> Action Committee on Access to Justice in Civil and Family Matters, "National Action Committee on Access to Justice in Civil & Family Matters: Prevention, Triage and Referral Working Group" (2013) Canadian Forum on Civil Justice at 3, online: *Osgoode Digital Commons* <https://digitalcommons.osgoode.yorku.ca/cfcj/60> [<https://perma.cc/YE6Q-K78Y>].

Rebecca Sandefur, a sociologist and access to justice scholar in the United States neatly articulated how these Everyday Legal Needs surveys provided a “radically different perspective”<sup>27</sup> from earlier forms of data in the justice sector. She described how the legal industry had historically constructed the access to justice problem as one that could be solved by adding more formal justice solutions. She pointed to the embedded assumption that all problems with a legal element required a legally trained professional. She disagreed with this assumption. She argued that research showed that people do not see their problems as legal problems, but rather as life problems: problems with their work, their housing, and their neighbors.<sup>28</sup> People often resolved these “everyday” problems outside the formal justice system, even though the problems contained a legal element. Sandefur advocated for improving access to justice by investing in research and building the evidence base that would provide information on opportunities to improve the formal justice system, and on other options for resolution outside the formal system.<sup>29</sup>

### The Data Deficit

Sandefur’s approach points directly to a weakness in access to justice work: the lack of evidence upon which any improvements may rely. When there is no evidence regarding how a system is doing, then decisions made in the hopes of improving the system can only be based on assumptions, estimates, and subjective opinions. Further, when there is no baseline data, then there is no way to assess if changes made have any lasting positive effect. Those working to improve access to justice also suggest that – at least in Canada - there is a lack of a tradition of

---

<sup>27</sup> Rebecca L Sandefur, *Access to What?* (2019) 148:1 Daedalus 49 (“Sandefur 2019”) at 50.

<sup>28</sup> Sandefur 2019, *supra* at 51.

<sup>29</sup> Sandefur 2019, *supra* at 53-54.

empirical research in law and that this has resulted in the justice system being “suspicious or fearful or both”<sup>30</sup> when exposed to these kinds of research methods and the resulting data.<sup>31</sup>

Whatever the reasons for the lack of data, Canada’s Roadmap recognized that making evidence-based decisions was a key part of improving access to justice and called for “reliable and meaningful metrics” which would support the evaluation of reform efforts.<sup>32</sup>

Remarking on a data deficit is not the same as saying that the system is without data. The formal justice system has always generated statistics regarding use and workload. For example, British Columbia’s court services collects data on the number of new court cases, court appearances, court sitting hours, and documents filed with the court. Beginning in 2018, Court services published this information on interactive, online “dashboards”.<sup>33</sup> This data is valuable for maintaining and improving the administration of justice. Other stakeholders in the formal system similarly collect data that supports their own departments. From this, we might presume that there is data available to support evidence-based decisions on where and how to make improvements.

A closer look, however, finds that this does not follow. The data described above is narrowly focussed on measuring the intake and output of the formal justice system. It does not provide information from a user perspective, for example, of a person’s experience within the system.

---

<sup>30</sup> Maurice Rosenberg, "The Impact of Procedure-Impact Studies in the Administration of Justice" (1988) 51 Law and Contemporary Problems 13 at 13.

<sup>31</sup> This is an interesting proposition, and one that intersects with theories of change and systemic change in the justice system. I will leave this to be considered in a further paper.

<sup>32</sup> Canada’s Roadmap, *supra* note 1 at 23.

<sup>33</sup> Ministry of Attorney General, “CSB Court Activity Dashboard”, online (webapp): < [© 2024 Katherine M. Gower](https://app.powerbi.com/view?r=eyJrIjoiYzY3YWU0NDMtODBmOS00YmQ4LWJkZTIhYTg5OTVhOGU1MDM1IiwidCI6IjZmZGI1MjAwLTNkMGQtNGE4YS1iMDM2LWQzNjg1ZTM1OWFkYyJ9> [https://perma.cc/52CY-52PV].</a></p></div><div data-bbox=)

Moreover, there is no systemic coordination of the data collected. Different stakeholders in the system often collect different categories of information. This makes comparisons across different datasets difficult or impossible.<sup>34</sup> Privacy concerns and statutory restrictions clog the channels where data could flow.<sup>35</sup> All of which means that there is a data deficit for those seeking to take an evidence-based approach to improving the formal justice system.

In the nascent shift to a person-centred approach to access to justice, the data deficit is even more acute. Taking a person-centred approach means recognizing that people manage their legal problems both inside and outside the formal justice system. Traditional methods of gathering administrative data are not constructed to gather information on problems outside the formal system. Indeed, one would be forgiven for wondering, “Where would we begin to gather the problems access to justice scholars and policy makers ought to be concerned about? How would we identify a legal problem if it has not crystalized by an interaction with the formal justice system?” Earlier, I introduced Everyday Legal Needs surveys. These surveys are carefully constructed to gather people’s experiences with legal problems which appear and are dealt with outside the formal system. The surveys ask after problems in different problem categories, for example, housing and employment. They use non-legal language to inquire how often people experience a problem in these categories which might have had a legal solution, or which raised legal issues such that a certain amount of legal information could have helped to secure a just resolution. Everyday Legal Needs surveys have become the dominant source of data supporting

---

<sup>34</sup> Jerry McHale Q.C. “‘The Justice Metrics Problem’ Background Paper” (March 3, 2017) Law Foundation of BC Grant No LLR3410 - 2016 UVIC ACE A2J Research Colloquium Project (unpublished) (“Metrics”) at 5.

<sup>35</sup> Metrics, *supra* at 6-7.

the person-centred approach to improving access to justice.<sup>36</sup> Each time a survey is undertaken, it returns a rich dataset for scholars and policy makers. However, Everyday Legal Needs surveys are undertaken only every two, three or even five years, and sometimes the resulting data is inaccessible for research (something I shall discuss more in this thesis). The result is that, while surveys do gather solid evidence to support person-centred work, there remains a paucity of evidence – a data deficit – which must be addressed to support the new approach to access to justice work.

In the next chapter, I provide a review of Everyday Legal Needs surveys in Canada and British Columbia. I begin with the landmark Everyday Legal Needs survey undertaken in the United Kingdom in the 1990s, and then describe the surveys' arrival in Canada and the changing focus and format of the Canadian surveys up to the most recent survey undertaken by Statistics Canada in 2021. I discuss the content of these surveys and their results and note particular characteristics and methodological assumptions.

---

<sup>36</sup> Trevor C.W. Farrow and Lesley A. Jacobs “Introduction: Taking Meaningful Access to Justice in Canada Seriously” in Trevor C.W. Farrow and Lesley A. Jacobs (Eds.) *The Justice Crisis: The Cost and Value of Accessing Law* (Vancouver: UBC Press, 2020) (“Farrow and Jacobs 2020”) at 11; Pleasence, P. 2016, “‘Legal Need’ and Legal Needs Surveys: A Background Paper” (London: Open Society Justice Initiative, 2016) at 4; Ab Currie, “The Legal Problems of Everyday Life: The Nature, Extent and Consequences of Justiciable Problems Experienced by Canadians” (Department of Justice Canada, 2007) (“Currie 2007”) at 3, online: <publications.gc.ca/pub?id=9.567261&sl=0> [https://perma.cc/QD3M-YHZE]. Internationally, work is underway to support the use of legal needs surveys to help countries understand and measure progress towards person-centred justice systems, see: OECD and Open Society Foundations, “Legal Needs Surveys and Access to Justice” (2019) online (pdf) *OECD* < https://www.oecd-ilibrary.org/docserver/g2g9a36c-en.pdf?expires=1708815962&id=id&accname=guest&checksum=1915929BA2D207365F1FDB46738D31AF> [https://perma.cc/R2XX-N9UU]. There is also qualitative legal research underway investigating people’s legal needs, but I have chosen to focus my thesis on quantitative data. For more information, Statistics Canada contracted community-based researchers to undertake quantitative work which would complement its 2021 Canadian Legal Problem Survey, online at: https://www.justice.gc.ca/eng/rp-pr/jr/survey-enquete.html [https://perma.cc/A5TS-6JS7].

In Chapter 3, I introduce a new incubator lab, the Justice Data and Design Lab (JDD Lab). The JDD Lab was conceived of as a lab that would engage with the new culture outlined in Canada's Roadmap. Like Everyday Needs Surveys, it would take a person-centred, evidence-based approach in its work.<sup>37</sup> But the JDD Lab's work would be fundamentally experimental in nature. The work would be undertaken in four-month terms, and after each term the Lab would share what it had learned. The full expression of the JDD Lab would be when data and data analysis could identify opportunities to improve access to justice and pass those evidence-based opportunities to prototyping teams. As will be discussed in this thesis, it is the experimental nature of the work which pushed the Lab to develop a new source of data on access to justice, to identify and apply initial tools to analyze that data, and to engage in this comparison of Everyday Legal Needs data and the new data. The Lab would also employ people from across multiple disciplines in all aspects of its work. When it encountered the challenges that inevitably crop up in research and development, it would draw from this interdisciplinary foundation to reorient its work. This last aspect proved invaluable as the Lab's work unfolded, making possible the Lab's pivot away from survey data and towards collecting and analyzing new data for access to justice from the social media platform, Reddit. The JDD Lab was positioned to serve as an action-oriented example of a methodologically-sound, user-centred, evidence-based, collaborative and experimental process. Chapter 3 describes where the Lab started, its pivot to Reddit data, and the resulting data collection and analysis.

---

<sup>37</sup> For a video overview of the Lab, see: Kate Gower, "Better Justice - How Data Scientists and Legal Scholars Are Collaborating to Help Canadians Solve Their Everyday Legal Problems" (October 2021), online (video): University of Victoria, Dean's Lecture Series <<https://continuingstudies.uvic.ca/humanities-and-social-sciences/series/deans-lecture-series/better-justice/>> [<https://perma.cc/KM4U-YKA5>].

Having reviewed the evolution of Everyday Legal Needs surveys in Canada and the data they offer, as well as the work undertaken at the JDD Lab and the data it collected and analyzed, Chapter 4 assesses the two data environments. The chapter discusses the strengths and weaknesses of the two datasets by juxtaposing the new Reddit data against the data from Everyday Legal Needs surveys. It concludes that Reddit data - subjected to the kinds of aggregation and analysis described in this thesis - can be a valuable, complementary source of data to that which is gathered in the surveys.

Chapter 5 gathers the lessons learned, including key takeaways for access to justice scholars navigating this information space. It also shares how the JDD Lab is being augmented and changed to explore new data, methods, and questions.

## My Role

This thesis summarizes work undertaken in the first six months of the JDD Lab's work, from January to June 2021. I was the lead researcher for the project, located at the University of Victoria's Faculty of Law. As lead researcher, I was responsible for project design, staffing, day to day oversight and decision-making, domain expertise in law and analysis of results brought forward by the data scientists on the team. I prepared Work Package Briefs ("Briefs")<sup>38</sup> to provide a framework for the data work.<sup>39</sup> The Briefs set out the objectives of the Lab, as well as a task list and a schedule for those tasks. It included timelines for regular reports to the Project

---

<sup>38</sup> Kate Gower, Better Justice Lab, "Division 2.0 – User Experience / Data Science: Work Package Briefs" dated December 11, 2021 ("Briefs") (unpublished) available upon request.

<sup>39</sup> Note that the Justice Data and Design Lab is made up of two interdependent parts. The first is a research engine which identifies evidence-based opportunities to improve access to justice. The second is a group of rapid response teams who take the identified opportunities and create prototypes to address those opportunities.

Board, who provided general oversight as well as financial responsibility. In the Briefs, I identified where I saw risks that the Lab would not produce the expected results (both by producing better than expected and less than expected results) and my thoughts on how to manage those risks.

The research question in the Briefs asked a broad question: “How can User Experience data and Data Science guide us on where to focus on improving access to justice in BC?” It deliberately created space for me to guide the team to think broadly about the issues and the data, and how the tools available could be brought to bear to answer the question. The Briefs provided enough infrastructure within which to do the work, but left space for decisions about the content of the work to be decided based on the data and people’s skills and abilities, under my scientific oversight.

As an access to justice champion who is trained in both the law and science, I was uniquely placed to design and execute this work. I was responsible for initiating all data science and legal scholars into the project. I introduced them to the concept of Access to Justice, and particularly to the Lab’s goal of supporting people who are resolving their legal problems outside of the formal justice system. I was responsible for constantly reaffirming and re-explaining the legal foundation of the project to keep the Lab within the ambit of the Briefs. I was also responsible for identifying moments when the timelines in the Briefs demanded final decisions, since further work depended on key decision points. As will be discussed, one key early decision concerned the very data the lab would use of people’s experiences of their legal problems. In the process of gathering the information upon which to make that decision, I had the opportunity to use all my

previous experiences in access to justice work, including my own relationships with people working in the field. I also took the opportunity to use my team building and leadership skills to create an environment in which the JDD Lab team could “cross-pollinate” all their skills and ideas.

The Lab successfully delivered all four of the results on the timeline identified in my Briefs. I prepared three of the results myself using both video and written reports. I guided and coordinated the interns so that they could learn from the experience of delivering the fourth report.

## CHAPTER 2: EVERYDAY LEGAL NEEDS SURVEYS AND DATA

In 1996, Professor Hazel Genn undertook a landmark Everyday Legal Needs survey in the UK, and reported the process and results in her 1999 book *“Paths to Justice: What People Do and Think About Going to Law”*.<sup>40</sup> She ran the first survey in England and Wales, and immediately undertook a second survey in Scotland.<sup>41</sup> Twenty years later, twenty-four large scale national surveys had used Genn’s methodology to learn more about the legal needs of their populations, including in surveys in Australia, Bulgaria, Canada, the Democratic Republic of Congo, Hong Kong, Northern Ireland, Japan, Moldova, the Netherlands, New Zealand, Taiwan, Slovakia, the Ukraine and the USA.<sup>42</sup>

Genn’s survey was commissioned by the Nuffield Foundation at a time when sweeping changes were underway in the British justice system, including reforms to civil procedure described in the Woolf Report and reconsiderations of public funding for legal advice. Policy makers and researchers found themselves with little data with which to assess the proposals for reform. Some data was available about people’s experiences in the formal justice system, but there was strong evidence that most people with a legal problem did not go to the formal justice system. There was a remarkable *lack* of information about what people were doing to resolve their problems outside the formal system. Genn’s work was designed to address the data gap. The survey was

---

<sup>40</sup> Genn 1999, *supra* note 7.

<sup>41</sup> Hazel Genn and Alan Paterson, *Paths to Justice: Scotland* (Scotland: Hart Publishing, 2001).

<sup>42</sup> Pascoe Pleasence, Nigel J. Balmer, and Rebecca L. Sandefur, “Paths to Justice: A Past, Present and Future Roadmap” (August 2013) at 5, online (pdf): *Nuffield Foundation* <<https://www.nuffieldfoundation.org/sites/default/files/files/PTJ%20Roadmap%20NUFFIELD%20Published.pdf>> [<https://perma.cc/MN24-K69Y>]; Nicole Aylwin and Mandi Gray, "Selected Annotated Bibliography of National and Regional Legal Needs Survey" (2015) *Canadian Forum on Civil Justice* 48, online (pdf): Canadian Forum on Civil Justice <https://digitalcommons.osgoode.yorku.ca/cgi/viewcontent.cgi?article=1048&context=cfcj> [<https://perma.cc/CL8E-XVKF>]

crafted to provide data which policy makers could rely on to make evidence-based decisions about how and where to help people, and assessments of the efficacy of any changes undertaken.

The great challenge Genn faced was how to shape a survey so it would gather information on the problems people faced which - while not resolved using the formal system - nevertheless contained a legal element. These were a subset of all problems, and not a subset that had received a lot of consideration. Genn and her team coined the term “justiciable event” to describe these problems, defined as:

“a matter experienced by a respondent [to Genn’s survey] which raised legal issues, whether or not it was recognized by the respondent as being “legal”, and whether or not any action taken by the respondent to deal with the event involved the use of any part of the civil justice system.”<sup>43</sup>

Genn and her team put significant time and energy into building a methodology to capture justiciable events in people’s lives. The methodology involved four phases. In the first phase, members of the research team, the public, the legal community and other stakeholders met and developed a “screening” questionnaire to gather information on how often people experienced justiciable events. Genn’s team carefully reviewed and discussed the format and language required. Questions were framed to gather matters which raised legal issues, without using legal terminology.

The survey set a threshold of importance which a person’s problem needed to exceed in order to be gathered by the survey. Genn set the threshold low. Respondents were asked to report all problems that were “difficult to solve”. They were only asked further questions regarding

---

<sup>43</sup> Genn 1999, *supra* note 7 at 12.

justiciable events if they had taken further action because they deemed the problem significant enough that it warranted their time.<sup>44</sup> In this way, Genn's survey collected information on what she called "non-trivial justiciable civil problems and disputes". The survey was also shaped to collect evidence on more than just the incidence of justiciable problems and the public's responses to them. It asked after people's motivation for taking action, any perceived barriers to access to justice, the outcomes of people's efforts, and people's experiences and perceptions of the process. The team ran pilot surveys of the screening questionnaire and the second, "main" questionnaire, to refine and perfect the surveys.<sup>45</sup>

With the surveys in hand, teams could begin gathering information. In the second phase, the team conducted interviews using the first screening questionnaire. The team interviewed people in person, at the person's home. The survey asked people to recall problems from the past five years relating to 14 different problem types. It asked after problems with employment, three different aspects of housing, faulty goods and services, money, family matters, matters respecting children, injuries, discrimination, unfair police treatment, immigration, and problems with neglectful medical treatment (for full details, see Appendix 1). Genn recognized the need to prompt people's memories, and interviewers used a set of "show cards" in the questioning.<sup>46</sup> Each problem type had an attendant "show card", which listed examples of that type of problem. For example, the show card for employment included the specific problem examples of:

- Losing a job (e.g. unfair dismissal...),
- Getting pay or a pension, other rights at work (e.g. maternity leave, sickness pay...),
- Changes to terms and conditions of employment,

---

<sup>44</sup> Genn phrases this in the negative: "... events that had been reported were excluded from further questioning if respondents said that they had taken no action whatsoever to deal with the problem because the problem had not been regarded as important enough to warrant any action." Genn 1999, *supra*.

<sup>45</sup> Genn 1999, *supra* at 16-17.

<sup>46</sup> Genn 1999, *supra* at 22.

- Unsatisfactory or dangerous working conditions,
- Harassment at work, and
- Unfair disciplinary procedures.<sup>47</sup>

Show cards were also used to prompt people's memories regarding if and what action they had taken in response to their problems. The list of possible actions was:

- Talked/wrote to the other side about problem,
- Sought advice about trying to solve problem,
- Threatened other side with legal action,
- Went to court- tribunal/arbitration, or started case,
- Went to mediation or conciliation,
- Took problem to an ombudsman,
- Took other action to solve the problem,
- Did nothing.<sup>48</sup>

In total, 4125 people completed the screening questionnaire.<sup>49</sup> The results identified a subset of 1134 people who had experienced at least one non-trivial justiciable event in the 5 years prior to the survey and taken action to resolve it.<sup>50</sup> In the third phase, researchers completed the main questionnaire with that subset of people. The fourth phase collected qualitative information from an even smaller subset of people.<sup>51</sup>

The results showed that 40% of people surveyed had experienced one or more justiciable event in the past five years.<sup>52</sup> The top seven most common problems were problems with faulty goods and services, money, injuries and health, owning residential property, living in rented property, relationships/family, and employment.<sup>53</sup> The study re-affirmed that most people do not go to the formal justice system to resolve their problems.<sup>54</sup>

---

<sup>47</sup> Genn 1999, *supra* at 22.

<sup>48</sup> Genn 1999, *supra* at 22.

<sup>49</sup> Genn 1999, *supra* at 16.

<sup>50</sup> Genn 1999, *supra* at 23.

<sup>51</sup> Genn 1999, *supra* at 16.

<sup>52</sup> Genn 1999, *supra* at 23.

<sup>53</sup> Genn 1999, *supra* at 23-24.

<sup>54</sup> Genn 1999, *supra* at 252.

The study also showed that many people experience the same problem more than once, and that many also experience more than one problem. It showed that certain problems occur in “clusters”. For example, people who experienced problems with employment also experienced problems with money, consumer issues, and problems to do with owning property. In another example, people who experienced divorce, also experienced problems with family, children, and/or money.<sup>55</sup> Genn commented on the omnipresent nature of problems with faulty goods and services.<sup>56</sup>

Genn emphasized that the type of problem was significant. Earlier work had focussed on learning more about the type of person experiencing the problems<sup>57</sup> while the problems themselves were considered as a homogenous background upon which different people struggled differently towards resolution. Genn argued that this was a flaw in previous research. Her work showed that certain types of problems were seen as important enough to warrant legal advice, and people took that approach (for example, problems with family breakdown and divorce). However, other types of problems were seen as less important and people decided those were problems that they could handle on their own (for example, problems with faulty goods). She wrote:

“The study reveals clearly the error in thinking about legal or other advice needs in relation to what certain types of “people” do, rather than focussing on what people do *in relation to particular problems*. The most self-reliant and confident people will, for certain types of problems almost automatically obtain legal advice, because the matter is important and because it is immediately characterized as a problem for which legal assistance is required.” (emphasis in original)<sup>58</sup>

---

<sup>55</sup> Genn 1999, *supra* at 31 and 249.

<sup>56</sup> Genn 1999, *supra* at 249.

<sup>57</sup> Genn 1999, *supra* at 6 and 253.

<sup>58</sup> Genn 1999, *supra* at 253.

Genn concluded that what people wanted was to resolve their problems, and to move on with living their lives. They want resolutions that were quick, cheap and relatively stress-free.

Genn's methodology arrived in Canada in 2004 when Ab Currie – who was then employed in the Research and Statistics Division of the Department of Justice, Canada – undertook the first of what would be three surveys on Canadians' experience with justiciable events.<sup>59</sup> Currie ran national surveys in Canada in 2004, 2006 and 2008.<sup>60</sup> His first survey focused on low- and moderate-income Canadians. The goal was to gather evidence on the need for support with legal problems within what he called “groups that are most vulnerable to problems”<sup>61</sup>. However, in 2006 Currie began collecting information from all Canadians.<sup>62</sup> The stated goals of the 2006 and 2008 surveys were to provide policymakers with evidence on both the incidence of legal problems in the population, and the extent of the unmet need for help resolving those problems.<sup>63</sup> Currie described the work as a “first step” that supported access to justice policy.<sup>64</sup>

Currie adapted Genn's methodology to fit the needs and resources available in Canada. The surveys were done by phone rather than in person. Currie considered the effects of these different delivery systems<sup>65</sup> and remarked that the reported incidence of problems was often higher for

---

<sup>59</sup> Ab Currie, “A National Survey of the Civil Justice Problems of Low- and Moderate-Income Canadians: Incidence and Patterns” (2006) 13:3 *International Journal of the Legal Profession*, 217-242 (“Currie 2006”).

<sup>60</sup> Ab Currie, “The Incidence of Justiciable Problems in Civil Matters in Canada: Three National Surveys in 2004, 2006 and 2008” (2009) Department of Justice Canada (“Currie 2009a”) online: CanLIIDocs <<https://canlii.ca/t/t90f>> [<https://perma.cc/V3QT-8HQ9>].

<sup>61</sup> Currie 2006, *supra* note 59 at 217.

<sup>62</sup> Currie 2007, *supra* note 36 at 3.

<sup>63</sup> Currie 2007, *supra* note 36 at 1.

<sup>64</sup> Currie 2009a, *supra* note 59 at 11.

<sup>65</sup> Ab Currie, “The Legal Problems of Everyday Life”, in Sandefur, R.L. (Ed.) *Access to Justice* (Sociology of Crime, Law and Deviance, Vol. 12) (Leeds: Emerald Group Publishing Limited, 2009) 1 (“Currie 2009b”) at 3.

telephone interviews, compared to in-person interviews. He mused whether people who responded in phone surveys were those more likely to have problems and wish to speak of them.<sup>66</sup> Currie's surveys also asked about people's experiences in the preceding three years rather than the preceding five years. Currie reviewed the growing international database of Everyday Legal Needs surveys and noted that across all surveys, the periods used for the reporting of problems differed, being set at either one year, three years, three and a half years, or five years. He felt there was no particular advantage to be gained by choosing any of these time frames.<sup>67</sup> However, he did note that the fact different surveys chose different time frames made it hard to compare results.

Refining the content of the surveys, Currie amalgamated the questions from Genn's screening questionnaire and main questionnaire into one survey. He shrank certain question categories and added new ones. For example, the 2004 survey collapsed Genn's three questions regarding property ("owning residential property", "renting rooms or property" and "living in rented property") into one section on "housing problems". It also added new questions on problems with "social/income assistance", "disability benefits", "wills and power of attorney" and "threats of legal action." The final 2004 Survey asked after a list of 15 problem types<sup>68</sup> where Genn's had asked after 13 problem-types (for full details, see Appendix 1).

Like Genn before him, Currie carefully shaped the survey questions. He asked respondents about their experience with the 15 problem types and, for each problem type, asked a series of

---

<sup>66</sup> Currie 2007, *supra* note 36 at Footnote 17.

<sup>67</sup> Currie 2007, *supra* note 36 at 11; Currie 2009a, *supra* note 60 at 6.

<sup>68</sup> Currie 2009a, *supra* note 60 at 4.

questions to prompt people's responses. These prompting questions were essentially the same questions which Genn's team had put on show cards. For example, when a respondent was asked about employment problems, the questions included generally the same questions Genn had asked, and also asked an additional question regarding getting unemployment insurance.<sup>69</sup> By the 2006 survey, prompting questions isolated information from 76 different specific "sub-problems" within the 15 categories.<sup>70</sup> In the 2008 survey, Currie added a new problem type - "Neighbourhood Problems" - which included sub-problems such as vandalism, property damage and noise. The 2008 survey collected information on 16 different problem types, and isolated information on 86 separate sub-problems.<sup>71</sup>

Currie also carefully considered the issue of "threshold language". In the 2006 survey, Currie set the threshold by asking respondents to report only problems that were "serious and difficult to resolve".<sup>72</sup> He also explored what this meant in people's lives. He included questions asking how important a problem was to resolve and how difficult a problem made a respondent's day to day life.<sup>73</sup> He found a person could have a serious problem which they might also say was either not important to resolve or did not make their lives much more difficult. He referred to this as the "elasticity" of serious and difficult problems.<sup>74</sup>

The 2006 survey also included strong wording that problems reported under one category were not to be reported under another. For example, the 2006 survey began with questions about

---

<sup>69</sup> Currie 2007, *supra* note 63 at 93.

<sup>70</sup> Currie 2009b, *supra* note 65 at 5.

<sup>71</sup> Currie 2009a, *supra* note 60 at 4.

<sup>72</sup> Currie 2007, *supra* note 36 at 32; Currie 2009a, *supra* note 60 at 5-6.

<sup>73</sup> Currie 2007, *supra* note 36 at 31-36; Currie 2009b, *supra* note 65 at 8-9.

<sup>74</sup> Currie 2007, *supra* at 32-36.

consumer problems. Later when questions were asked about debt problems, respondents were instructed to only provide information on problems “not related to any problems already mentioned”<sup>75</sup>.

In total, 4501 low- and middle-income Canadians completed the 2004 survey, 6665 completed the 2006 survey and 7002 completed the 2008 survey.

Currie’s three surveys consistently reported that about half of Canadians experience a justiciable event in any three-year period.<sup>76</sup> The surveys confirmed, again, that most people handle their problems on their own, outside of the formal justice system.<sup>77</sup> Consumer, debt and employment problems were always reported as the top problems people experienced. With the addition of neighbourhood problems in 2008, the top problems became consumer, neighborhood, debt and employment in that order.<sup>78</sup> Currie remarked that, by introducing questions regarding neighbourhood problems in the 2008 Survey, the percentage of people reporting a serious and difficult to resolve problem jumped from 49.9% to 54.6%. He remarked that a survey will return a higher percentage of people with problems as the number of problem types queried increases.<sup>79</sup>

Currie used a standard cluster analysis to investigate how people’s problems occurred in clusters.<sup>80</sup> The cluster analysis used an algorithm which began by imagining each problem type as one cluster, and then continued to combine clusters until only one cluster remained. Currie

---

<sup>75</sup> Currie 2007, *supra* at 93.

<sup>76</sup> Currie 2009b, *supra* note 65 at 7.

<sup>77</sup> Currie 2007, *supra* note 36 at 55 and 60.

<sup>78</sup> Currie 2009a, *supra* note 60 at 9.

<sup>79</sup> Currie 2009a, *supra* at 7.

<sup>80</sup> Currie 2009b, *supra* note 65 at 23; Currie used the methodology described in Pascoe Pleasence, *Causes of Action: Civil Law and Social Justice*, 2nd Ed (Legal Services Commission (UK), 2006) at 65-72.

displayed the results on “dendrograms”, also known as “tree diagrams”<sup>81</sup>. Only the data for people who reported five or more legal problems generated distinct clusters.<sup>82</sup> Currie found similar clusters to those Genn had found before him. For example, he found a distinct cluster of debt, employment, and consumer problems. However, having included new problem categories, he also identified new clusters. For example, he identified a cluster of immigration, disability pensions and social assistance problems - these last two being problem types which he had added to the survey.

Currie remarked, as other researchers before him had done, that the hypothesis that problems cluster together because they share a causal link is not proven by the cluster analysis. Rather, clusters could occur because of issues of social vulnerability or other circumstances in people’s lives. To investigate how problems were connected, his surveys overtly asked respondents if they felt certain problem types had “triggered” other problems. He combined the clustering and triggering data and generated five clusters in his 2006 data. He reported three of them. From largest to smallest, they were: first, a cluster of problems triggered by relationship breakdown problems, which included debt, threat of legal action, other family law problems and consumer problems; second, a cluster of problems triggered by consumer problems, which included legal action and debt problems; and third, a cluster of problems triggered by personal injury problems, which included employment and debt problems.<sup>83</sup>

---

<sup>81</sup> Currie 2009b, *supra* note 65 at 22-25.

<sup>82</sup> Currie 2009b, *supra* at 29.

<sup>83</sup> Currie 2009b, *supra* at 28.

Currie's surveys were robust. They gathered information on the demographics of respondents and on what people did in response to their problems. The surveys also collected information about the costs associated with problems, the physical and mental effects of the problem on Canadian's lives, and people's perceptions of the justice system.

By 2013, Currie had joined forces with Trevor Farrow at the Canadian Forum on Civil Justice ("CFCJ"). With a small team of researchers, they undertook a fourth survey of everyday legal needs in Canada titled "Everyday Legal Problems and the Cost of Justice in Canada" (the "CFCJ Survey")<sup>84</sup>.

The CFCJ Survey generally followed Genn's methodology, as adapted by Currie. It was done by telephone. The CFCJ Survey continued the threshold methodology of asking for information only on "SERIOUS problems that were not easy to fix" (emphasis in survey provided to interviewers).<sup>85</sup> It collected demographic data. It added a new problem category - "criminal charges" - and collected information on 17 different problem types and 84 sub-problems.<sup>86</sup> It began with new questions about a respondent's general attitude about the justice system.<sup>87</sup> It also wove in a new layer of questions, hoping to gather information on the cost of people's everyday legal problems. This was something Currie had started in his earlier surveys. The CFCJ took an expanded view of "cost", using it to include time lost, lost opportunities, and the effect on

---

<sup>84</sup> David Northrup, Ab Currie, Trevor C.W. Farrow, Les Jacobs, and Nicole Aylwin, "Everyday Legal Problems and the Cost of Justice in Canada: Survey" (2016) Canadian Forum on Civil Justice, Toronto, Canada ("2013/2014 Survey") at 2, online (pdf): *Canadian Forum on Civil Justice* <<https://www.cfcj-fcjc.org/sites/default/files/Everyday%20Legal%20Problems%20and%20the%20Cost%20of%20Justice%20in%20Canada%20-%20Survey.pdf>> [<https://perma.cc/AYX6-YES8>].

<sup>85</sup> 2013/2014 Survey, *supra* at 2.

<sup>86</sup> 2013/2014 Survey, *supra* at 3-14.

<sup>87</sup> 2013/2014 Survey, *supra* at 2.

people’s physical and mental health i.e. cost in the form of a loss of health.<sup>88</sup> The CFCJ team conducted the survey over several months in 2013 and 2014.<sup>89</sup> Researchers performed telephone interviews with over 3000 adults across Canada<sup>90</sup>.

The CFCJ Survey reaffirmed most people deal with their problems outside the formal system, and that the top problem types in Canada were consumer, debt, employment, and neighbourhood. It showed that just under 50% of survey respondents experienced a problem in a three-year period. Nearly a third of them reported experiencing more than one problem.<sup>91</sup> The data showed that nearly a third of those people did not resolve their problem(s) in the three-year period.<sup>92</sup> 95% of respondents took steps to try and resolve their problem(s).<sup>93</sup> Of those who took action, 75% attempted to negotiate with the other party, 61% asked for advice from friends and relatives and 33% went to the Internet looking for help.<sup>94</sup>

Seven years later, Statistics Canada conducted Canada’s most recent survey of everyday legal problems and disputes (the “StatsCan Survey”).<sup>95</sup> The StatsCan Survey was part of

---

<sup>88</sup> For more on this see: Ab Currie “The Monetary Costs of Everyday Legal Problems and Expanding Access to Justice” in Trevor C.W. Farrow and Lesley A. Jacobs (Eds.) *The Justice Crisis: The Cost and Value of Accessing Law* (Vancouver: UBC Press, 2020) (“Currie 2020”) at 93.

<sup>89</sup> Trevor C.W. Farrow, Ab Currie, Nicole Aylwin, Les Jacobs, David Northrup and Lisa Moore, “Everyday Legal Problems and the Cost of Justice in Canada: Overview Report (Canadian Forum on Civil Justice)” (Toronto, Canada, 2016) (“Farrow et al 2016”), online (pdf): CFCJ <[https://www.cfcj-fcjc.org/sites/default/files/Everyday Legal Problems and the Cost of Justice in Canada - Overview Report.pdf](https://www.cfcj-fcjc.org/sites/default/files/Everyday%20Legal%20Problems%20and%20the%20Cost%20of%20Justice%20in%20Canada%20-%20Overview%20Report.pdf)> [<https://perma.cc/D3CF-76EP>].

<sup>90</sup> Researchers interviewed 3051 adults via land line phone, and a further 212 via cellphone. (Farrow et al 2016, *supra* at 4, footnote 9.

<sup>91</sup> The CFCJ Survey found 48.4% of Canadians experienced a legal problem in a three-year period and 30% of those people reported that their problems were not resolved and were ongoing (Farrow et al 2016, *supra* at 6-7).

<sup>92</sup> Farrow et al 2016, *supra* at 11.

<sup>93</sup> Farrow et al 2016, *supra* at 9.

<sup>94</sup> Farrow et al 2016, *supra* at 9.

<sup>95</sup> Canada – Statistics Canada, “Canadian Legal Problems Survey - Detailed information for 2021” (January 18, 2022) (“StatsCan Survey”), online: *Statistics Canada* <<https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5337>> [<https://perma.cc/87CC-VAPD>]; Laura Savage and Susan McDonald, *Experiences of serious problems or disputes in the Canadian provinces, 2021* (January 18, 2022) Juristat, Statistics Canada (“Savage and McDonald 2022”), online (pdf): *Statistics Canada*

the federal government’s response to the United Nations Sustainable Development Goals, particularly Development Goal 16, which included providing “equal access to justice for all”.<sup>96</sup> The survey was conducted online from February to April 2021.

The StatsCan Survey asked respondents to report disputes in 18 problem categories which generally matched the categories from the CFCJ survey<sup>97</sup>; however, the survey shifted the methodology with respect to the prompting questions. Genn’s methodology, followed by Currie and Farrow, asked prompting questions as each problem type was broached. In the StatsCan Survey<sup>98</sup>, when respondents were first asked to identify if they had had a problem in the 18 problem types being canvassed, their answers were limited to “yes” or “no”. There were no prompting questions. Then the survey went through the problem types again, asking if the problem(s) experienced in each problem type were “serious and not easy to fix”. Again, the answers were limited to “yes” and “no”, and there were no prompting questions. A short series of questions regarding the impact of the Covid-19 pandemic followed, after which, the survey returned to the problem types. This time, the survey asked prompting questions which are similar to those used by Genn, Currie and Farrow. However, the survey only returned to a problem type if the respondent had answered “Yes” to having a problem of that type

---

<<https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2022001/article/00001-eng.pdf?st=93IAg3mu>> [<https://perma.cc/DXL7-HHJP>].

<sup>96</sup> Savage and McDonald 2022, *supra* at 4; UN Goal 16, *supra* note 4. The UN’s Strategic Development Goals were released in 2015. More information is available online: <https://sdgs.un.org/goals#history> [<https://perma.cc/MASC-B8U>].

<sup>97</sup> The StatsCan Survey gathered disputes in 18 categories rather than 17 because “debt” was changed from one to two categories by asking after disputes with respect to “Your house, your rent, your mortgage or rent owed to you”, and “Money you owe (debt) or money owed to you, excluding rent and housing issues.” (emphasis added) StatsCan Survey, *supra* note 95.

<sup>98</sup> StatsCan Survey, *supra*.

when first taken through the 18 problem type questions. Finally, after a series of questions regarding how problems could be connected to one another, the StatsCan survey asked the respondent to identify their “most serious problem” from the problems they had already reported. The survey collected demographic data about the people answering the survey (including visible minorities, people with disabilities and Indigenous peoples).

The StatsCan Survey reported that 34% of Canadians surveyed experienced a problem in the previous three years, and 18% of respondents experienced a “serious”, “not easy to fix” justiciable event.<sup>99</sup> These numbers are lower than previous surveys, something I will discuss further in Chapter 4. The report that accompanied the survey provided information on the “most common problems” and on the “most serious problem” respondents reported (Table 1).

---

<sup>99</sup> Savage and McDonald 2022, *supra* note 85 at 5.

<b>Most Common Problem Canada</b>	<b>Most Serious Problem Canada</b>	<b>Most Serious Problem BC</b>
Neighbourhood: Vandalism, property damage and noise (21%)	Neighbourhood: Vandalism, property damage, noise (10.85%)	Neighbourhood: Vandalism, property damage, noise (12.79%)
Receiving poor or incorrect medical treatment (16%)	Consumer: Large purchase or service (9.11%)	Poor or incorrect medical treatment (9.15%)
Being discriminated against (16%)	Poor or incorrect medical treatment (9.00%)	Employment: Employer or job (8.19%)
Being harassed (16%)	Employment: Employer or job (7.77%)	A will, or taking care of financial issues for another (7.82%)
Having a problem with a large purchase or service (15%).	Debt or money owed to you (7.69%) <sup>100</sup>	Health (7.58%), Debt/money owed to you (7.41%), Consumer: Large Purchase (7.36%)

**Table 1: Most Common and Most Serious Problems in Canada and BC in 2021**

*Source: StatsCan Survey*

In the same time frame in which Currie ran his three surveys across Canada, Legal Aid British Columbia (“LABC”) started conducting its own series of surveys in that province. Given its mandate, LABC’s surveys focused on the types of problems low-income British Columbians faced on an everyday basis. The surveys’ goal was to help understand better how low-income British Columbians managed their problems, and how the problems affected their lives and their perceptions of the justice system. LABC’s surveys did not include questions on how problems related to each other or how they clustered. They were administered online.<sup>101</sup> LABC’s surveys ran on a 5-year interval, in 2008, 2013 and 2018. In 2018, LABC updated its survey. It noted that comparisons between the first two surveys and those from 2018 and beyond should be made with

<sup>100</sup> After this, there is a drop to Harassment (7.19%) and Discrimination (6.34%).

<sup>101</sup> Legal Aid BC - Everyday Legal Problems (August 24, 2020) (“LABC 2020”) at 5, online (pdf): *Legal Aid BC – Reports* <[https://legalaid.bc.ca/sites/default/files/2020-09/Everyday%20Legal%20Needs%20Survey\\_1.pdf](https://legalaid.bc.ca/sites/default/files/2020-09/Everyday%20Legal%20Needs%20Survey_1.pdf)> [<https://perma.cc/73AF-73GR>].

caution.<sup>102</sup> LABC ran an “additional” survey in 2020 to collect information regarding the impact of the Covid-19 pandemic on the everyday legal problems of low-income British Columbians.

The surveys generally used Genn’s methodology. In 2018 and 2020, the surveys asked after 13 problem types.<sup>103</sup> Like Currie, LABC collapsed Genn’s three categories regarding housing into one, and added “welfare and social assistance”, “wills and power of attorney”, “police incidents” and “threat of legal action”. There was no “neighbourhood problems” category. As a threshold, respondents were asked to report problems that were “serious and difficult to resolve”.

Prompting questions were asked as each problem type was broached.<sup>104</sup> The surveys collected demographic information.

The 2020 survey found that 83% of low-income British Columbians experienced at least one of the 13 legal problem types in 3 years. This was slightly up from 76% in 2018. This echoed Ab Currie’s early findings in 2004 which indicated low- and middle-income people experienced a higher number of legal problems than the general population. The top problems reported in the 2020 survey were consumer, employment, money or debt, housing or land, and discrimination.<sup>105</sup>

This brings us to the end of the overview of Everyday Legal Needs surveys undertaken in Canada and British Columbia. Though I will engage in a thorough review of the strengths and

---

<sup>102</sup> Legal Services Society - Everyday Legal Problems (July 19, 2018) (“LABC 2018”) at 4, online (pdf): *Legal Aid BC – Reports* <[https://legalaid.bc.ca/sites/default/files/2019-03/lssEverydayLegalProblems07\\_2018.pdf](https://legalaid.bc.ca/sites/default/files/2019-03/lssEverydayLegalProblems07_2018.pdf)> [<https://perma.cc/TE3Y-TASY>].

<sup>103</sup> Consumer, Employment, Money or debt, Housing and Land, Personal injury, Discrimination, Family relationship, Welfare or social assistance, Wills and Powers of Attorney, Hospital treatment or release, Police incidents, Legal action, Immigration. LABC 2020, *supra* note 101 at 16.

<sup>104</sup> LABC 2018, *supra* note 102 at 3; LABC 2020, *supra* note 101.

<sup>105</sup> LABC 2020, *supra*.

weakness of these surveys in Chapter 4, there are three characteristics and methodological assumptions which I shall remark on here. First, these surveys take a person-centred approach. They gather data from the perspective of the people who experience legal problems, rather than from an institution which has collected administrative data regarding the number and frequency of problems it addresses. The surveys rely on Genn's definition of a "justiciable event" to gather information on events people experience which contain a legal element, whether or not those people engage with the formal justice system to manage them. They do not require respondents to have any knowledge of the substantive law related to civil legal problems. For example, there is no requirement for respondents to identify their problems as legal, nor are respondents required to have taken any action to resolve their problems using the civil justice system. The surveys, and the data they generate, have allowed scholars and policy makers to learn about the prevalence and characteristics of civil legal problems impacting the public in their day-to-day lives.

Second, these surveys assume that asking questions about pre-set problem types will provide accurate information on the problems people face. I am particularly intrigued by this assumption, given that the list of problem types was regularly adjusted as each new survey was undertaken. Genn used 13 problem types. Currie adapted Genn's methodology to use in Canada. When he did so, he condensed Genn's three problem types concerning housing into one, and he added problem types which considered social income assistance, disability benefit, wills and power of attorney, and the threat of legal action. Currie ultimately asked after 15 problem types. In 2006, Currie added "neighbourhood problems". In 2013, the CFCJ Survey added "criminal charges". In 2021, the StatsCan Survey made a different kind of adjustment to the problem types. I

particularly note that it restructured questions on family problems to be two problem types of “Family Breakdown, Excluding Child Custody” and “Child Custody”, where earlier Canadian surveys had structured family problems to be “Family Problems: Relationship Breakdown, Breakup, Divorce, Child Custody” and “Other Family Problems”. This list of changes to the problem types in different surveys raises interesting questions: What would happen if we used different problem types from those which are used now? How do we know we are using the “best” list of problem types? What would happen if we used no problem types?

Third, it may seem trite to say so, but Everyday Legal Needs surveys are surveys. Whether they engage the public in person, by telephone or via the internet, surveys ask respondents to recall experiences and report them after they have occurred. There is a methodological assumption that the surveys will nevertheless adequately gather evidence on people’s experiences. I am not saying that these surveys do not return accurate results, rather I am making more visible the inherent risk in asking people to recall events when asked to do so, at a time well-removed from when the event occurred. The surveys themselves include features that both acknowledge and minimize this risk. All the surveys, except for the StatsCan Survey, used prompting questions to support the data collection as each problem type was addressed. For example, Genn’s “show cards” regarding employment problems prompted respondents with sub-problems such losing a job, having trouble securing pay or pension, or experiencing unsafe working conditions. Currie added another prompt, or “sub-problem”, when he asked after problems with getting unemployment insurance. By the CFCJ survey, respondents were prompted with questions about 84 different sub-problems.

I close this chapter by noting that, while this review has brought us up to date with the Everyday Legal Needs surveys undertaken in Canada and British Columbia, there have only been five nation-wide surveys and the LABC surveys, though valuable, focus only on low-income British Columbians. More data is needed to build a person-centred evidence base of people's everyday legal needs before we can reasonably say we have overcome the current data deficit. The next chapter introduces a new lab which uncovered a new and different type of person-centred data. The chapter tells the story of how the JDD Lab uncovered a source of person-centred data which is not built around a constructed definition of a justiciable problem, and which does not impose problem types on the people providing the data.

### CHAPTER 3: THE JUSTICE DATA AND DESIGN LAB (the JDD Lab) and REDDIT DATA

The JDD Lab was designed to collect and analyze data of British Columbians' legal problems to identify where people were asking for more assistance resolving their problems outside of the formal justice system. Like the Everyday Legal Needs surveys, the JDD Lab was constructed to take a person-centred approach, and collect data from the public's perspective of their experiences with legal problems. However, the JDD Lab planned to do something different with the data. It planned to subject the data to the tools available and regularly used in disciplines other than law, beginning with applied data science. In this way, the Lab would take a fresh look at the data and, hopefully, provide new insights into the problems people experienced and the help they were asking for. As noted before, the JDD Lab began its work intending to use the data from Everyday Legal Needs surveys, but - for reasons that are described below - quickly pivoted to gathering a new kind of data. The Lab identified the data by investigating where people looked online for help with their legal problems, and found that the majority of people looking online for help use the social media site, Reddit.

This thesis reports on the data collection and analysis undertaken in the first six months of the JDD Lab's work, from January to June 2021. The overarching methodology at the JDD Lab was to collect, assemble, and analyze data and report the results. Those results would provide information to answer the JDD Lab's research question: "How can User Experience data and Data Science guide us on where to focus on improving access to justice in BC?"

## Collection

The first two tasks in the Work Package Briefs (“Briefs”) were to form the interdisciplinary team and to assemble the first datasets. The Briefs were based on the assumption that certain Everyday Legal Needs datasets would be available. In particular, the Briefs assumed the JDD Lab would have access to the data collected by Trevor Farrow and the team at the Canadian Forum on Civil Justice in 2013/2014 (the “CFCJ Dataset”)<sup>106</sup> and the 2004, 2006 and 2008 datasets gathered by Ab Currie (“Currie’s Dataset”) (collectively, the “Datasets”).

Best efforts were made to collect the Datasets. I engaged with all levels of both the JDD Lab and the CFCJ. However, by February 1, 2021, I determined that this data would not be available on the Lab’s timeline. The Datasets appeared to be held behind administrative and privacy requirements which could not be overcome in the Lab’s timeline. In terms of the Lab’s work, I found myself in the unenviable position of not having the data which appeared to have been built into the methodology. However, I reframed the problem so that it was not the Lab that had failed, but rather a key assumption that had failed. When that assumption failed, the Lab was squarely in the data gap discussed in Chapter 1. At the time, the assumption that the Datasets would be available appeared to be a fundamental element of the research. Without the Datasets, it momentarily felt as if the project were at an end. In fact, when that assumption failed, it became the catalyst that changed what we saw and how we thought about data.

We carefully reviewed the data we had. Evidence from the CFCJ Dataset showed that the majority of people did not go to the formal justice system to resolve their problems, and that 33%

---

<sup>106</sup> Farrow et al 2016, *supra* note 84.

of those people went to the internet for help. Working in the spring of 2021, one year into the Covid-19 pandemic, we were reasonably confident that this percentage would have increased, and we set off to learn where people who were looking for help went online.<sup>107</sup>

We used existing online analytic programs to explore where people went online looking for help. We learned that when people searched on Google for “legal advice”, the top place people in BC went was to the social media platform Reddit. We used Google Trends<sup>108</sup> to explore interest over time in the term “legal advice”. Google Trends provided “Related topics” and “Related queries” which showed the other topics and searches “googled” by people who searched for “legal advice”<sup>109</sup> (see Figure 1). Since the JDD Lab was focussed on BC, we looked past the top result that referenced resources in Ontario, and focussed on the second results, both of which mentioned Reddit. On SEMRUSH, we researched “legal advice” using SEMRUSH’s “Keyword Magic Tool”<sup>110</sup>. As with Google Trends, SEMRUSH listed “Reddit” as the second related keyword for which people searched, after an Ontario-related search (see Figure 2). Using this information, the team located the subreddit “r/legaladvicecanada”.<sup>111</sup>

---

<sup>107</sup> The JDD Lab was right. The 2021 StatsCan Survey showed that the top two places where people look for help with their most serious problem are family and friends and the internet (both being used by 51% of respondents who looked for help outside the formal justice system) (Savage and McDonald 2022, *supra* note 95).

<sup>108</sup> Google Trends, online: <<https://trends.google.com/trends/>> [<https://perma.cc/7UXX-KF74>]. Google Trends is a free service that allows people to see what people are searching for in Canada online and in real time.

<sup>109</sup> Google, “Find Related Searched”, online: <https://support.google.com/trends/answer/4355000> [<https://perma.cc/K4SV-6G7B>].

<sup>110</sup> Semrush is subscription-based, online product. It can be used to track searches for particular keywords and to reveal what other related searches people undertake when looking for those keywords. I note, however, that Semrush markets itself as a marketing tool, rather than a research tool. Semrush, online: <<https://www.semrush.com>> [<https://perma.cc/R8FM-RL4C>].

<sup>111</sup> Further research on where people look online for legal help is underway at the JDD Lab.

Related topics <span>?</span>		Rising <span>▼</span> <span>⬇️</span> <span>⏪</span> <span>⏩</span>
1	Legal Aid Ontario - Corporation	Breakout
2	Reddit - Website	Breakout <span>⋮</span>
3	Family law - Topic	Breakout <span>Explore Reddit - Website</span>
4	Court - Topic	Breakout
5	Contract - Topic	Breakout

< Showing 1-5 of 11 topics >

Related queries <span>?</span>		Rising <span>▼</span> <span>⬇️</span> <span>⏪</span> <span>⏩</span>
1	free lawyer advice	Breakout
2	legal advice reddit	Breakout
3	lawyers	Breakout
4	free legal aid	Breakout
5	free family law advice	Breakout

< Showing 1-5 of 10 queries >

**Figure 1: Google Trends results for “legal advice” (Spring 2021)**

Keyword	Volume	Trend	KD %	CPC \$	Com.	SERP Fe...	Results
free legal advice ontario	1,600		71.72	1.54	0.70	+2	56.8M
reddit legal advice	1,300		81.37	3.00	0.02	No	972M
free legal advice	1,000		75.09	1.38	0.72	+2	1.5B
legal advice	1,000		78.02	5.61	0.47	+3	1.6B
free legal advice bc	720		66.78	1.36	0.64	+3	52.2M
free legal advice alberta	480		65.67	2.40	0.82	+3	14.7M
legal advice ontario	480		70.98	5.24	0.81	+3	83.3M
free legal advice calgary	390		59.62	1.99	0.80	+3	8.6M
24 hour free legal advice ontario	320		65.46	1.89	0.83	+3	8.6M
free legal advice toronto	320		75.69	2.28	0.79	+2	21.8M

**Figure 2: SEMRUSH “Keyword Magic Tool” results for “legal advice” (Spring 2021)<sup>112</sup>**

<sup>112</sup> SEMRUSH, “Keyword Magic Tool”, online: <<https://www.semrush.com/kb/262-keyword-magic-tool>> [<https://perma.cc/2HV9-L5QN>].

## Reddit

Reddit is a social media platform that allows people to post stories and questions, and have other people on Reddit provide comments and information in response.<sup>113</sup> Users (known as “Redditors”) can choose to join conversations on particular topics (known as “subreddits”). From 2019 to 2021, Reddit had the third greatest growth in users, behind only TikTok and Pinterest, increasing its number of monthly active users by over 30%.<sup>114</sup> In 2021, the Pew Research Centre in the USA found Reddit was one of the very few social media platforms in the USA whose user group was growing.<sup>115</sup> In 2022, Redditors worldwide posted 430+ million posts to Reddit.<sup>116</sup> Reddit use is also growing in Canada. According to a 2022 report from the Social Media Lab at Ryerson University, 19% of Canadian adults who are online have a Reddit account, a number which has been steadily increasing from 9% in 2017 and 15% in 2020.<sup>117</sup>

The subreddit “r/legaladvicecanada” was created in December 2015. The subreddit is titled “Legal Advice for Canadians” and described as “[a] place to ask simple legal questions.”<sup>118</sup> The subreddit has rules for posting and commenting, including a list of things which shall be considered “off topic” and removed. These include personal advice, anecdotes, and comments

---

<sup>113</sup> Reddit, “About”, online: *Reddit* <<https://www.Redditinc.com>> [<https://perma.cc/4UZ5-9XQR>].

<sup>114</sup> Statista, “Growth of Monthly Active Users of Selected Social Media Platforms 2019-2021”, Online: *Statista* <<https://www.statista.com/statistics/1219318/social-media-platforms-growth-of-mau-worldwide/>> [<https://perma.cc/F7Y7-R58D>].

<sup>115</sup> Brooke Auxier and Monica Anderson, “Social Media Use in 2021” (April 2021), online: *Pew Research Centre* <<https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>> [<https://perma.cc/8PPB-4MJZ>].

<sup>116</sup> Reddit, “Revealing This Year’s Reddit Recap, Where We Highlight How Redditors Kept It Real in 2022” (December 8, 2022), online: *Reddit* <<https://www.Redditinc.com/blog/Reddit-recap-2022-global:>> [<https://perma.cc/6ATF-AG7M>].

<sup>117</sup> Philip Mai and Anatoliy Gruzd, “The State of Social Media in Canada 2022: A Census-Balanced Survey About Social Media Adoption and Use in Canada” (Toronto: Social Media Lab Toronto Metropolitan University, 2022) at 16.

<sup>118</sup> r/legaladvicecanada, *Legal Advice for Canadians*, online: r/legaladvicecanada <<https://www.Reddit.com/r/legaladvicecanada/>> [<https://perma.cc/FY2W-PKPK>]. The subreddit emphasizes advice provided is for information purposes only and should not be taken as “final or official advice”, concluding: “See a local lawyer for the best answer to your questions”.

not containing a legal answer. Like other subreddits, “r/legaladvicecanada” has a panel of moderators who ensure posts comply with the rules.

Posts are personal stories in which people provide some context of their situation, and generally ask for advice regarding their legal rights. People provide a title, and then post their story in their own words. The posts are first-person descriptions of their experience and their problem. The posts represent a unique opportunity to hear the stories of real people who find themselves facing what they either know or suspect is a legal problem. From the JDD Lab’s point of view, when people post publicly on Reddit looking for help, their posts are potentially a valuable dataset. The team was hopeful that a dataset, created out of Reddit posts, could be analyzed to show what problems were dominating the subreddit. We were now in a position to use aggregation methods and analytical techniques familiar to the data scientists, and to see what insights our interdisciplinary team could reveal using this data.

### Reddit Data Collection and Aggregation

The Team pulled a first set of posts from the subreddit “r/legaladvicecanada” and began a subjective review which suggested the subreddit would be an excellent source of data. People were joining “r/legaladvicecanada” in droves seeking legal help. Membership in the subreddit went from 42,000 members in March 2021, to 86,000 members in June 2021.<sup>119</sup>

---

<sup>119</sup> This Reddit forum continues to grow. In November 2021, membership was 187,000 members; in November 2022, membership was 409,000; and in August 2023, membership was 443,000. Reddit, *Legal Advice for Canadians*, online (subreddit) *Reddit* <<https://www.Reddit.com/r/legaladvicecanada/>> [<https://perma.cc/XDT7-S8GM>].

Since the JDD Lab’s goal was to gather data regarding the needs of people in British Columbia, we took steps to limit the datapulls to collect only posts concerning BC legal problems. In Canada, Reddit directs Redditors to post the province in which they reside and we wrote computer code that used this location requirement to gather BC-specific posts. Fears that this would shrink the dataset too much for it to be useful proved to be unfounded. The datapulls, filtered to include only BC-produced posts from Reddit, brought back 790 posts in April 2021 which grew to 1987 in June 2021. The number of words to be analyzed from all posts reached 468,000 by June 2021.

Early on, the team conducted a semi-automatic clustering analysis on the first 790 posts using “lists” in Python. Python is a computer programming language whose source code is available to the public, allowing for changes and development as users see fit. It is easy to use – compared to other programming languages – and it has certain advantages that make it particularly good for the JDD Lab. For example, it can process large volumes of data and has a large library of pre-written code (again, available publicly) that allows programmers to perform complicated mathematical, and statistical analysis, as well as to access and use machine learning and artificial intelligence with relative ease. At this early stage, “Lists” analyzed the titles and content of all posts and allocated them to set groups. The result was distinct clusters, with Employment, Housing and Family problems appearing as the most common problems.

By June 2021, the Team had constructed a web application to display datapull results and analysis completed using machine learning programs. The application was created using Django, another open-source program which makes it easier to build web applications and websites. Data

collection programs ran when a team member manually clicked a button titled “Update Data Lake with latest data” (password required). The update process connected to Reddit’s application programming interface (Reddit’s “API”), fetched the latest posts from the subreddit, and uploaded the data collected as “json”<sup>120</sup> files to a Reddit responses “bucket” online. “Json” files are a standard way of sharing data so that the structure one person has added to the data, or created to hold the data, is transmitted to the next person who uses that data.

For a database management system, the team built a PostgreSQL database with scalable computing and storage buckets in Google Cloud. Programming and notes were shared on Github<sup>121</sup>, a website which acts as an online platform where developers can collaboratively write programming code. Data were timestamped with the time and date of the data aggregation. Any information that could identify a Redditor was removed.<sup>122</sup> Data could be exported for detailed review by researchers (for example, in an excel spreadsheet).

PostgreSQL (pronounced post-GRES-q-l) is another open-source product. It is a strong “object-relational” database system, something I will discuss more below. The Lab used cloud computing (a “virtual” machine) to run the database management system, “renting” time on a per second basis. The intent was (and is) to give easy access to future team members, and those interested in the Lab and its data. It is also scalable. Early costs for the Lab’s work were low. Future plans include having the Lab “scale up” as the project develops.<sup>123</sup>

---

<sup>120</sup> “json” stands for “JavaScript Object Notation”.

<sup>121</sup> Github keeps track of any changes made to the code and who made them.

<sup>122</sup> This data is housed in separate online storage and is available should it be needed to advance further analysis.

<sup>123</sup> There is undoubtedly more to be written on cloud computing, but that topic is outside the scope of this paper.

The scalable repository combined with the relational database management system holds data in what is called a “data lake”. Other approaches would have required all collected data to fit into a database with a predetermined structure. Instead, the data lake can - and currently does – hold both structured and unstructured datasets in their native format. The intention is that, in the future, the JDD Lab can collect and aggregate data from many different data sources, both structured and unstructured, and offer the possibility of “cross-pollinating” different datasets, which were previously considered incompatible with one another.<sup>124</sup> Holding data in its native format also means not imposing pre-set categories on data. This allows for data analysis which finds patterns of categories contained in the data itself. An example is Latent Dirichlet Allocation, an unsupervised machine-learning clustering program which will be discussed more below.

## Analysis

The early subjective review and semi-automated analysis of Reddit posts gave the JDD Lab team confidence that Reddit posts contained valuable information about people’s legal needs for data analysis. Subjectively, we concluded the posts were rich with data. Other than providing a title to head the post, people’s posts were unstructured, free-flowing descriptions of their personal experience. The posts held information about what problems people were facing and the significance of that problem in their lives (see Figure 3, below). Some Reddit posts suggested the Redditor understood some of the legal elements in their problem. Other posts simply described a person’s experience, and then asked, “is this legal?” Furthermore, the semi-automatic

---

<sup>124</sup> The power and utility of data lakes and relational databases are topics for further work. For more see: B. Douglas Blansit “The Basics of Relational Databases Using MySQL” (2006) 3:3 *Journal of Electronic Resources in Medical Libraries*, 135; Britta Kruger, “Why You Should Use a Relational Database Instead of a Spreadsheet” (2004) 35:7-8 *Cybernetics and Systems: An International Journal* 683.

analysis gave the team confidence that clustering using unsupervised machine learning would provide usable returns.

*“Hello, I am moving. My landlord has refused to return security deposits to every tenant in the 2 years I have been here. I will clean my unit as specified by the Landlord Tenant Board, however, I'm concerned about “normal” wear and tear. The window blinds in the front are bent and the window lock is broken. I notified my landlord but he hasn't fixed anything in all the time I've been here. Do I need to replace the window blinds and lock or are these considered normal wear and tear? Should I say nothing and let arbitration decide?”*

**Figure 3: An Example Post from r/legaladvicecanada (edited for confidentiality)**

The team used Latent Dirichlet Allocation (“LDA”) to cluster the data into topics, and an inter-topic distance mapping program to create an interactive visualization of the topics. Both were available in Python. Data cleaning and preparation were written into the code.

LDA is a topic modelling program that, without supervision, runs through text identifying patterns of word usage and allocating terms into clusters. With LDA, a “topic” is a cluster of terms (i.e., words or phrases) which – once the program has reviewed the body of data - regularly appear together. This suggests shared subject matter. In this way, LDA may be said to be “discovering” topics that are “latent” in the dataset.

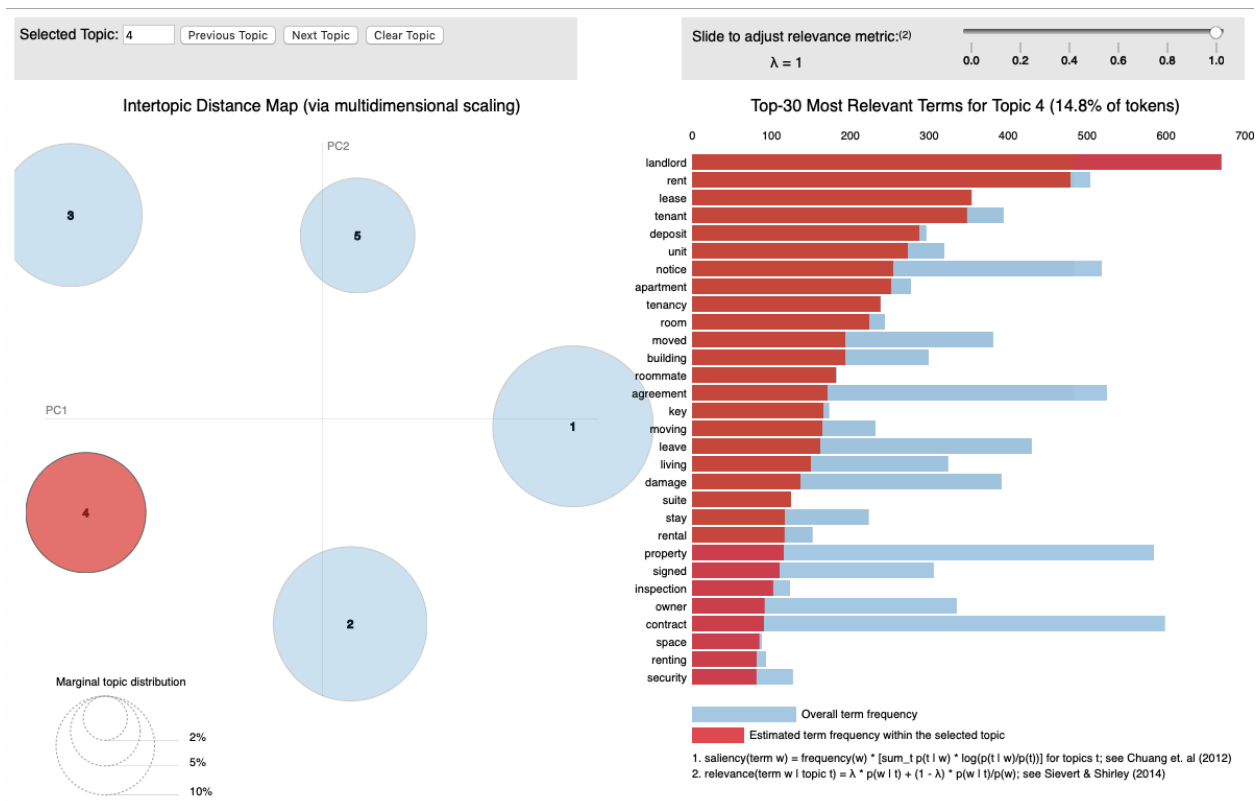
Once LDA had created clusters, inter-topic distance mapping plotted them across multiple dimensions. Then it scaled the information into something the human mind can comprehend which in this case was two dimensions.<sup>125</sup> The program generated an interactive window that

---

<sup>125</sup> “Getting to the Point with Topic Modeling | Part 3 - Interpreting the Visualization” (2020), online: *Data Science: Machine Learning and Data Science for Beginners and Experts Alike* <<https://community.alteryx.com/t5/Data-Science/Getting-to-the-Point-with-Topic-Modeling-Part-3-Interpreting-the/ba-p/614992>> [<https://perma.cc/BSG9-Q8TR>].

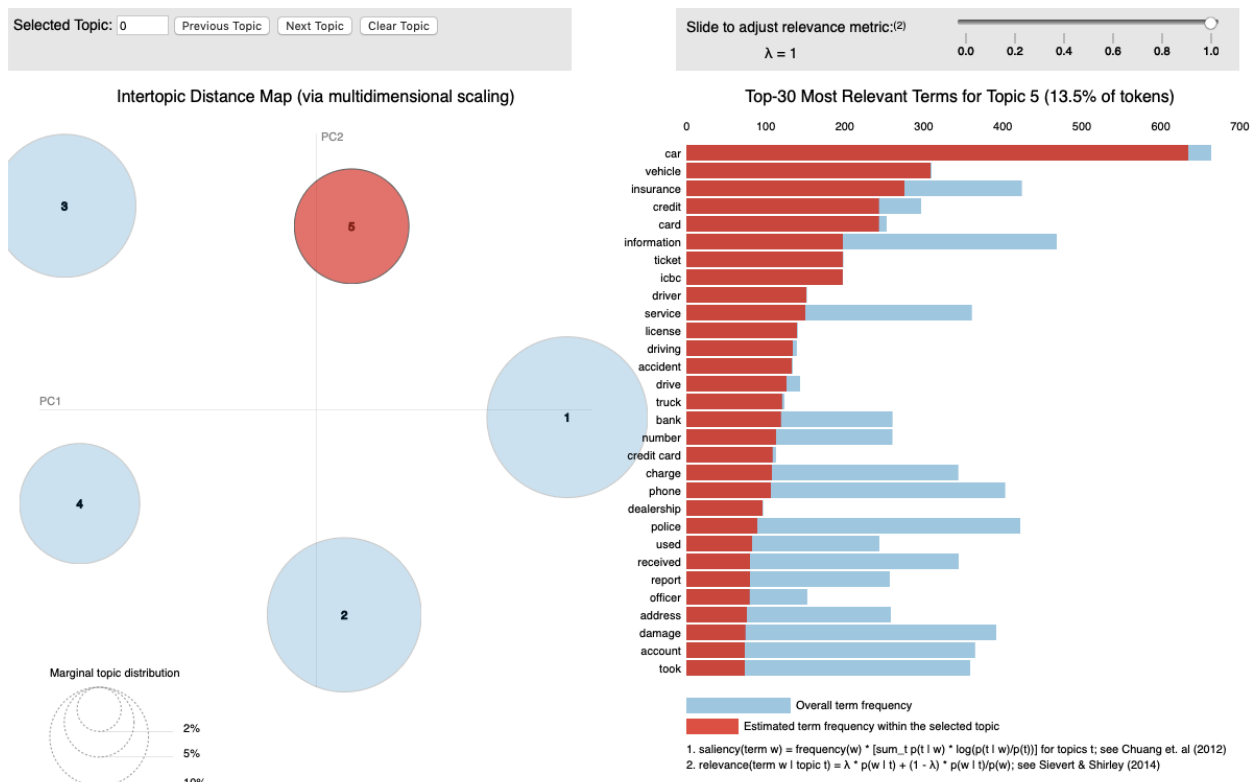
displayed red circles representing each cluster, and - alongside the circles - a list of the top 30 most “relevant”<sup>126</sup> terms in each cluster.

Figures 4 and 5 show the results of these two tools analyzing the Reddit data in the JDD Lab’s repositories as of June 23, 2021. The LDA has allocated the data into five clusters. The program itself suggested this was the “best” number of clusters for this dataset. The inter-topic distance map provides the visual of five red circles, each representing a cluster.



**Figure 4: Cluster 4 - Results of LDA and Inter-topic Distance Mapping on Reddit Data**

<sup>126</sup> “Relevance” in inter-topic distance mapping is a technical term defined by the program.



**Figure 5: Cluster 5 - Results of LDA and Inter-topic Distance Mapping on Reddit Data**

## Results

In Figures 4 and 5, the left side of the screen shows five circles representing five clusters or potential topic areas. The right side lists the top 30 “most relevant” terms. With inter-topic distance mapping, the further apart the clusters are, the more distinct they are from one another. The smaller a cluster is, the less word distribution there is, suggesting more cohesiveness and connectedness between the words in the cluster. In Figures 4 and 5, all five clusters are separate and distinct. Clusters four and five are the smallest and thus the most cohesive.

In Figure 4, Cluster 4 is highlighted, and the five most relevant terms are “landlord”, “rent”, “lease”, “tenant”, and “deposit”. Clicking on each term will show visually if that term only

appears in a single cluster (for example, “landlord” only appears in Cluster 4) or if the term appears in more than one cluster (for example, “tenant” appears in Clusters 4 and 1). The LDA analysis allows terms to appear in more than one cluster. This is known as “fuzzy clustering”.

These analytic tools identify words which the algorithm determines regularly appear together, but the algorithm knows nothing about what the topic might be. My legal domain expertise, and knowledge of local and world events, allowed the JDD Lab to identify topics and flesh out what kind of information was in a cluster. For example, while the top five words in Cluster 4 suggested the cluster concerned housing issues, legal expertise added the valuable insight that in British Columbia the Residential Tenancy Act applies to most rental housing situations, setting out many of the responsibilities of, and protections for, landlords and tenants.<sup>127</sup> There is also a residential tenancy branch of government that helps people navigate the tenant-landlord relationship, and organizations that provide help and support for both tenants and landlords.<sup>128</sup> This suggests that the data collected may contain information on specific opportunities to improve people’s experience with this Act, these processes and these organizations.

After a visual review of the top five terms in each cluster, the general topics in Clusters 5, 4, 3 and 2 appeared to be: car and insurance issues, housing issues, employment issues, and family-related issues. A visual review of the top five terms associated with Cluster 1 did not immediately suggest a topic (see Table 2: Top Five words in Clusters and Possible Topics).

---

<sup>127</sup> Residential Tenancy Act, SBC 2002, c 78.

<sup>128</sup> “BC Residential Tenancy Branch”, online: *Government of BC* <https://www2.gov.bc.ca/gov/content/housing-tenancy/residential-tenancies> <[<https://perma.cc/U4KN-RNKK>]; Tenant Resource and Advisory Centre, online: *Tenant Resource and Advisory Centre* <<https://tenants.bc.ca>> [<https://perma.cc/F4GT-D9J5>]; Landlord BC, online: *Landlord BC* <<https://landlordbc.ca>> [<https://perma.cc/5NHJ-RCDC>].

Cluster	Top Five Terms	Possible Topic
5	car, vehicle, insurance, credit, card	Cars and Insurance
4	landlord, rent, lease, tenant, deposit	Housing
3	contract, employer, manager, employee, employment	Work
2	child, parent, mom, father, school;	Family
1	property, dog, police, cost, lot.	[No Topic Identified]

**Table 2: Top Five words in Clusters and Possible Topics**

For the JDD Lab, Cluster 5 was the most exciting result. The top five terms were “car”, “vehicle”, “insurance”, “credit”, and “card”, and the second five were “information”, “ticket”, “ICBC”, “driver” and “service” (see Figure 5). The acronym “ICBC” stands for “Insurance Corporation of British Columbia”. In BC, vehicles on public streets must be insured by ICBC, and claims following accidents run through that corporation. On May 1, 2021, ICBC launched a new “no-fault” insurance model.<sup>129</sup> JDD Lab Reddit data was collected in April, May and June of 2021. I will systematically address the strengths and weaknesses of Reddit data in the next section, but it seems possible there is a correlation between Cluster 5 and this shift in ICBC’s policy. This suggests that regular data pulls in our data collection could show particular legal concerns that are in the consciousness of British Columbians in real-time.

---

<sup>129</sup> “New ICBC No-fault Insurance Model Launches in May” CTV News (Thursday, April 29, 2021) Online: *CTV News* <https://vancouverisland.ctvnews.ca/new-icbc-no-fault-insurance-model-launches-in-may-1.5408332> [<https://perma.cc/RC62-KA9K>].

## CHAPTER 4: ASSESSMENT

In the previous two chapters, I reviewed Everyday Legal Needs surveys and the data produced in Canada, and what the JDD Lab’s discovered when it collected and analyzed Reddit data regarding the legal problems people posted to that social media site. Now I turn to assessing the strengths and weaknesses of the two datasets by juxtaposing them against one another.

### Everyday Legal Needs Data

Everyday Legal Needs surveys have been the backbone of data supporting access to justice work since Hazel Genn’s first survey in 1996. Without the surveys, we would not know that nearly half of Canadians face a legal problem in any given three years. Nor would we know that the top justiciable events which they face are problems of vandalism, property damage and noise in their neighborhood, problems with purchasing goods and services, problems getting wages paid, unemployment insurance and other problems with work, and problems stemming from debt (i.e. the problem types “neighborhood”, “consumer”, “employment” and “debt”). After the most recent survey from Statistics Canada, we can add challenges with medical treatment to the list of top justiciable events.

We also know that these problems are serious for the people who experience them, that they affect a person’s health and well-being, and that the cost of resolving them is high (for both the persons immediately affected and society as a whole). Everyday Legal Needs surveys have shaded in features on what was previously the “unknown landscape” of civil problems. The surveys have given us evidence regarding where people look for help to resolve their problems. They have revealed a whole network of people, community groups and services upon which

people rely in their attempts to achieve resolution. Survey data has also taught researchers about the very nature of justiciable events, including how they multiply, trigger one another, and manifest as “clusters” of problems in a person’s life.<sup>130</sup>

These surveys changed how we saw the justice system. When Currie ran his first national survey in 2004, it was the first empirical evidence in Canada of what legal problems people experience outside the formal justice system.<sup>131</sup> They gathered empirical evidence from the users of the system and allowed us to see what the system looks and feels like from the users’ perspective. Currie explained this well when he wrote that Everyday Legal Needs surveys are “a way to give voice to the public.”<sup>132</sup> Like Genn before him, Currie saw the value of the data for helping policymakers to make evidence-based decisions. He described his first three surveys and the data they produced as “the first step in support of access to justice policy.”<sup>133</sup>

Currie’s early surveys supported Canada’s shift to taking a person-centred perspective on the justice system. When the National Committee prepared Canada’s Roadmap and advocated for a “people-first” approach, it described the “breadth and depth of problems in people’s everyday lives” using survey results from Genn and Currie. That survey data meant the National Committee could include evidence-based statements such as: “many people have everyday legal problems”, “the poor and the vulnerable are particularly prone to legal problems”, “problems multiply”, and “legal problems have social and economic costs”.<sup>134</sup>

---

<sup>130</sup> Currie 2009b, *supra* note 65 at 20-28.

<sup>131</sup> Currie 2007, *supra* note 36 at 1.

<sup>132</sup> Currie 2007, *supra* note 36.

<sup>133</sup> Currie 2009b, *supra* note 65 at 11.

<sup>134</sup> Canada’s Roadmap, *supra* note 1 at 2-3, citing Currie 2007, *supra* note 36.

These points in Canada’s Roadmap emphasized the particular value of the demographic information gathered in the surveys. That information allowed researchers to conduct valuable analysis of the experiences and actions of certain groups of people. Currie used demographic data to make multiple findings. He showed how everyday legal problems related to age, gender, language, and marital status.<sup>135</sup> He considered “predictors” for justiciable problems. He found disability was a significant predictor in all 15 problem categories<sup>136</sup> and he went on to consider key predictors for all problem types. For example, in debt problems the data indicated that having a disability and simply being a younger person were the top two predictors, and that - of the two - the more powerful predictor was that a person was young.<sup>137</sup> In another example, he found that the top predictors for experiencing a consumer problem were having a disability, being young and having three or more children.<sup>138</sup>

The most recent survey from Statistics Canada gathered information on respondents who were from visible minorities, who were people with disabilities and who were Indigenous. As a result, analysts could report conclusions about certain “person-types”. For example: “First Nations people, Métis and Inuit living in [Canada’s] provinces were much more likely than non-Indigenous people to report experiencing one or more serious problems in the three years preceding the survey (27% versus 18%)”<sup>139</sup>, “People with disabilities are much more likely than people without disabilities to experience three or more serious problems” and “Indigenous

---

<sup>135</sup> Currie 2007, *supra* note 36 at 19-25.

<sup>136</sup> Currie 2007, *supra* note 36 at 26.

<sup>137</sup> Currie 2007, *supra* note 36 at 27.

<sup>138</sup> Currie 2007, *supra* note 36 at 26-27 and 61.

<sup>139</sup> Savage and MacDonald 2022, *supra* note 95 at 6.

people are more likely to experience harassment and discrimination”<sup>140</sup>. This kind of data provides guidance for improving the justice system. When data shows that Indigenous people are more likely to experience a problem than non-Indigenous people and are also more likely to experience harassment and discrimination, then these are two data points that could be used to support work to improve those experiences.

Surveys have also regularly been constructed to gather information on topics of particular interest. During the Covid-19 pandemic, LABC added a one-time survey to collect information on how Covid-19 was creating and impacting the legal problems of low-income British Columbians. It ran the survey at a 2-year interval rather than the usual 5-year interval. The LABC 2020 survey reported that for low-income British Columbias, Covid-19 caused or worsened 36% of their legal problems. Further, for those with an unresolved problem, 45% said Covid-19 was affecting their ability to resolve the problem.<sup>141</sup> The CFCJ Survey was crafted to address what it called a “major gap” in the data regarding the impact of the economic and social, and individual and collective costs associated with everyday legal problems. The book that followed the survey reported the data, and also provided analyses, commentary and thinking to support evidence-based policy-making and reform.<sup>142</sup> The CFCJ survey results showed just over a quarter of people spent money on a legal problem, and that the overall cost “to the Canadian public” was \$7.8 billion dollars each year.<sup>143</sup> The book dedicated a chapter to considering if and how these high costs to both individuals and the knock-on effects to Canada’s institutions could impact policy decisions on spending.

---

<sup>140</sup> Savage and MacDonald 2022, *supra* note 95 at 7.

<sup>141</sup> LABC 2020, *supra* note 101 at 8.

<sup>142</sup> Farrow and Jacobs 2020, *supra* note 36 at 15.

<sup>143</sup> Currie 2020, *supra* note 88 at 95.

The person-centered data generated by Everyday Legal Needs surveys is invaluable evidence for those who work on justice policy and social policy generally. However, these surveys take significant time and money to prepare and execute. To prepare her survey, Genn and her team first worked with stakeholders to identify the goals of the survey, and then dedicated an entire stage of work to creating a survey that met those goals. As part of that first stage, her team carefully created questions to draw out people’s legal problems without using legal terms and ran pilot surveys to test the questions before finalizing the survey. This time and effort to craft a survey is essential. Without this work, a lower quality of survey may be sent out to collect data, resulting in poor data that may or may not meet the goals of the survey. Funding is required to support this early work, just as funding is needed to run the survey.

When Genn’s team administered that first survey, they conducted 4125 in-person interviews at people’s homes with their screening questionnaire and 1134 with the main questionnaire. Currie and Farrow’s surveys were done by telephone, connecting with 4501, 6665, 7002 and 3263 people for the 2004, 2006, 2008 and 2013-14 surveys respectively. If we conservatively estimate 30 minutes for an interview, interviewing the 3263 people in 2013 and 2014 would have taken nearly 68 days, working 24 hours a day.<sup>144</sup> This is without travel time, since the survey was conducted over the telephone. Genn’s team would have allocated time (and money) to travel as well.

---

<sup>144</sup> Statistics Canada may have improved on the investment required to run a national survey by administering its 2021 survey via an online portal. The portal was open for seven months in 2021. We shall have to wait and see how the dataset develops. LABC has successfully administered its 2018 and 2020 surveys online. LABC 2018, *supra* note 102; LABC 2020, *supra* note 101 at 5.

These costs and commitment are significant, and the result is that Everyday Legal Needs surveys are few in number. At best, they are conducted on a 3 to 5-year schedule. In Canada, there are notable gaps in the data since there was no survey between 2008 and 2013, and no survey between 2014 and 2021. Further, survey data is usually not available immediately to researchers once the survey is completed, as time is needed to process the data and prepare a summary of results. For example, the StatsCan Survey was undertaken in 2021, and was reported out in 2022, accompanied by a paper interpreting the results.<sup>145</sup>

Further, researchers must also take a close look at a survey's methodology before comparing survey results from different years. LABC carefully noted that it made updates for the 2018 survey including changes to the survey questions and qualifying criteria, and remarked that results from both 2018 and 2020 are “not directly comparable to prior waves”<sup>146</sup> but that those earlier surveys may still be used for general, broader comparisons.<sup>147</sup> At the national level, each new survey has made changes to the problem types queried. The most recent Everyday Legal Needs survey – the StatsCan Survey – went one step further, and used a new methodology for prompting respondents with sub-problems. The new methodology meant respondents were not prompted with sub-problems they might have faced in each problem type. Rather, they were asked about sub-problems later in the survey, and only if they had previously indicated that they had experienced that overarching problem type. In the StatsCan survey results, the incidence of problems was lower than in earlier surveys. It seems likely that when the StatsCan Survey asked respondents if they had experienced a problem, but did not offer any prompting questions, the

---

<sup>145</sup> Savage and MacDonald 2022, *supra* note 95.

<sup>146</sup> LABC 2020, *supra* note 101 at 4.

<sup>147</sup> LABC 2018, *supra* note 102 at 4.

survey collected a lower incidence of justiciable problems than may exist. As an example for consumer problems, a respondent to the StatsCan survey was asked if they had had any problems with “a large purchase or service where [they] did not get what you paid for and the seller did not fix the problem”. However, a respondent to the CFCJ Survey was asked this first question and then prompted with five further questions about sub-problems as follows: problems with paying for repairs or renovations or similar done to home, car or appliance; problems with a service such as to do with moving, club membership, holidays, tax preparation; problems with the safety of a purchase; problems with insurance claims being unfairly rejected; and/or problems with a bill or invoice.

Changes to survey format are to be expected as new teams take responsibility for national projects. It may be necessary to adjust the focus of a survey or to incorporate new theories on survey methodology. However, these changes mean that it may not be possible to compare the updated survey with previous surveys, and this robs researchers of a dataset which could be used to spot trends and assess impacts. In an ideal world, researchers would have access to surveys which were administered on a short interval and which used the same methodology. A series of these surveys would create a dataset upon which researchers could rely to identify patterns in people’s experiences or to assess the impact of new initiatives to improve access to justice. In Canada, researchers are gaining better access to the data from Everyday Legal Needs surveys, but a dataset for assessing patterns and the impact of new initiatives is not yet available.

Keeping all these points in mind, I turn to assess the Reddit data.

## Reddit Data

The Reddit data's immediate point of value is that it is a steadily growing dataset, and the method of posting information to the subreddit has remained the same since the subreddit was created in 2015. It gathers information daily, at little or no cost to researchers.<sup>148</sup> It is building a dataset that can be used to identify patterns in what help people are asking for as they work to resolve their problems. It fills in an obvious gap which the surveys are not filling. I am not suggesting that because of this, Reddit data is better than current survey data. In this section I will consider the strengths and weaknesses of this new data set. However, the very fact that data is accruing constantly makes it a valuable data source worth learning more about.

The JDD Lab's work confirms that Reddit data can provide valuable insights into people's everyday legal problems. Reddit posts responded well to the LDA clustering, and inter-topic distance mapping, generating five distinct clusters. Clusters five, four and three appeared as the smallest circles on the inter-topic display, indicating tightly connected collections of posts. Those clusters centered around vehicles and insurance, housing and employment, in that order. These are early results from the Lab, but remarkable ones. I cannot help but note, with some excitement, that these tightly connected clusters are different from the top issues generated by Everyday Legal Needs surveys, which have consistently been neighbourhood, consumer,

---

<sup>148</sup> At the time of writing, there is some public discussion regarding whether Reddit data will remain generally available in this way, however there is always discussion about ensuring the data will still be available for academic work. See subreddit "r/AskAcademia" online: <  
[https://www.Reddit.com/r/AskAcademia/comments/146w9sv/for\\_those\\_who\\_are\\_using\\_Reddit\\_data\\_for\\_research/](https://www.Reddit.com/r/AskAcademia/comments/146w9sv/for_those_who_are_using_Reddit_data_for_research/)  
> [<https://perma.cc/PDG6-LMKR>]. I also note that while research in Medicine using Reddit data is consistently found not to require ethical approval from University Ethic's boards, scholars remind us that the public-private divide is a blurry line when dealing with social media, we must be careful to ensure ethical considerations are always part of our research projects. See: S Ravn, A Barnwell, B Barbosa Neves. "What Is "Publicly Available Data"? Exploring Blurred Public-Private Boundaries and Ethical Practices Through a Case Study on Instagram" (2020) 15:1-2 *Journal of Empirical Research on Human Research Ethics* 40.

employment and debt. While acknowledging that there is much more work to be done, these initial results indicate a whole new world of data.

### Assessment

I offer five further strengths regarding this new data. First, the data is valuable as we pivot to a person-centred approach to improving the justice system. Each Reddit post is an example of a person working on their problem and asking for the help they want and need. Posts are unprompted, created by people who are motivated to resolve what is troubling them. There is little structure. People post when they wish and write what they wish. In the survey-based methodology, a researcher, external to the person's life, attempts to pull from that person information relating to their experiences. The respondent is asked to cast their mind back over the last three years and recall the details of justiciable events they have experienced. Compare this to Reddit, where there is no researcher asking a person to recall problems based on the questions in a survey. Rather, people choose to share information about the problems they have and the help they need. This information is coming from the "ground up", inductively informing researchers how to help people. This is the epitome of "person-centred" information.

To buttress this point, consider how Everyday Legal Needs surveys create a threshold in an effort to gather only "serious and difficult to solve" problems. Genn put considerable thought into designing a threshold to exclude "trivial problems"<sup>149</sup>. Currie discusses the subjectiveness of the threshold, since people are asked in the surveys to make the determination themselves regarding

---

<sup>149</sup> Genn 1999, *supra* note 7 at 13.

if problems are “serious”.<sup>150</sup> On Reddit, there is no language setting a threshold. A person decides when a problem is important enough to warrant them taking the time to find the subreddit and type out the details. Only those problems which, from a person’s perspective, are significant enough to warrant posting, make it into the Reddit dataset. The threshold is built into the mechanism. This threshold is still subjective, but it is different than what has been constructed in surveys. This means new information comes forward in the data.

It is also worth considering the effect of how, on Reddit, there are no prescribed problem types. Problem types are still there, “latent” in the data: the LDA and inter-topic distance mapping returned four identifiable clusters within the dataset. But it is particularly notable that this dataset returned three clusters that generally match the problem types in surveys (housing, employment and family) and one cluster with a problem type which has never been part of a survey (cars and insurance). Reddit data and analysis appears to capture unplanned-for (and unimagined?) problem types. Again, this is new information which we do not get from surveys.

Looking closer at the Reddit clusters, Cluster 5 was the most cohesive cluster and the top 10 relevant terms related to cars, insurance and ICBC (the vehicle insurance provider in BC). The data was collected in April, May and June 2021. In the middle of that collection period - on May 1, 2021 - ICBC launched a new, “no fault” insurance model. While this thesis considers the early datapulls collected by the JDD Lab, it is useful to note that later datapulls indicate that the cluster regarding cars and insurance no longer exists. It seems possible there is a correlation between this cluster and the shift in car insurance policy. We imagine that the shift to no-fault insurance

---

<sup>150</sup> Currie 2009b, *supra* note 65 at 8.

causes concerns. Survey evidence tells us that people look for help online. The Lab's analysis tells us that a top place where people look for help online is on Reddit. When people post to Reddit they post as they experience a concern about the policy change, and the data and LDA analysis generates a cluster of posts. In this example, Cluster 5 is generated as the policy change approaches and is brought into force, and then the cluster disappears.

This brings me to my second point about Reddit data: there is an "immediacy" woven through it, which is novel and valuable. At the JDD Lab, the Reddit dataset grows with each datapull and each datapull holds a "snapshot" of the issues people were posting on Reddit, represented as clusters. Considering clusters against current events and policy changes could provide completely new information from that which is gathered in surveys.

Medical researchers have demonstrated this phenomenon in Reddit data. In 2020, a team in the USA used Reddit data to consider how people with health anxiety fared during the Covid-19 pandemic.<sup>151</sup> The research team downloaded data from 28 subreddits: 15 that focused on specific areas of mental health, two that discussed mental health more broadly and 11 that were not health related. They analyzed the resulting dataset using topic modelling with LDA, and other Natural Language Processing techniques. The analysis showed patterns in the subreddits concerning mental health. A key result graphed a weighted count of Covid-19 related posts against news announcements from NBC news and found that posts in the subreddit

---

<sup>151</sup> D. M. Low, L. Rumker, T. Talker, J. Torous, G. Cecchi, and S. S. Ghosh, "Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit during Covid-19: Observational Study" (2020) 22:10 Med Internet Res e22635 12 ("Low et al. 2020").

“r/healthanxiety” spiked a full two months before other subreddits.<sup>152</sup> The authors posited that Reddit data and analysis could reveal patterns in users’ health anxiety which could help provide better support and resources to users - for example, revealing certain information earlier than other data sources.<sup>153</sup> If the JDD Lab can identify problems as they bubble to the surface in Reddit, then the medical research suggests that this puts us in a position to provide people better support and resources in a timely manner.

Third, let us consider the effect of people reporting their problems as they experience them, and because they are motivated to solve them. The details of their experience are fresh in their minds. The facts of their story are as accurate as they are able to make them. Reddit’s lack of formatting constraints allows a person to report as many facts as they can. This is a new form of “self-reporting”. This is an opportunity to analyze the further information contained in a free-flowing narrative of a person’s problem. In the next chapter I will outline the further work the JDD Lab plans to do with this information. For the moment I will “seed” the idea that there is data *within the data* in first person accounts such as these posts. A good example would be the sentiment – the feelings and emotion - that accompany the text.

Fourth, I cannot understate the value of the Reddit data simply being available. The story of the JDD Lab’s first days emphasizes how important that is. Recall that one of the Lab’s key assumptions failed (i.e., the anticipated survey data was not available on the Lab’s timeline). The

---

<sup>152</sup> Low et al. 2020 *supra* at 12, and MultiMedia Appendix S1. Other online data has the same potential. In 2017, the Pew Research Centre in the USA reviewed google search data and tracked it alongside events and news coverage. Ke Matsa, Amy Mitchel and Galen Stocking “Searching for News: The Flint Water Crisis” Pew Research Centre (April 2017) online: *Pew Research* <<https://www.pewresearch.org/journalism/2017/04/27/searching-for-news-the-flint-water-crisis/>> [<https://perma.cc/NA8Y-SW6R>].

<sup>153</sup> Low et al. 2020, *supra* note 151 at 14.

Datasets were held behind administrative and privacy requirements which could not be overcome on the JDD Lab's timeline. These sorts of barriers to accessing data are not uncommon, and they are obvious barriers to research. Statistics Canada has gone some way to changing this by making the recent StatsCan Survey dataset widely available. However, even though it was enormously disheartening, accepting that the Datasets were unavailable became the catalyst that changed what we saw as, and how we thought about, data. The team realized that the project only experienced a data deficit if we limited ourselves to the structured datasets which we had expected from Everyday Legal Needs surveys. Instead, we realized that the internet is full of data. Whenever a person writes online something such as "I have a problem with my family/credit card/rental. What can I do?", then that is data. The data are "unstructured": they are written in text, and not housed in the rows and columns of a traditional database. To anyone untrained in data management, this unstructured data appears inaccessible and unusable. However, data science techniques mean that unstructured data is at least as accessible and valuable as structured data, and perhaps more so for the reasons discussed above. Our explorations led us to Reddit, but the real importance of this moment is that the Lab went from wallowing in a "data gap", to being "data rich". The JDD Lab has plans to make this data available to other researchers in the field, something I will discuss more in the next chapter.

Fifth, the data analysis used to date does not require researchers to pre-specify problem types. Rather, LDA analyzes the actual language used in the source data and identifies problem clusters within the data itself. This means the results are less vulnerable to changes in the methodology which were discussed with respect to survey data. Instead, it is possible to run the cluster analysis at regular intervals and watch as clusters appear and disappear in the Reddit data. At this

point, the JDD Lab has only the Reddit data to review. However, the infrastructure of the Lab means data from other social media platforms can be added to the Lab’s data lake and included in the clustering analysis. The JDD Lab has plans to gather data from other social media platforms currently available, such as Facebook. Further, while Reddit is certainly a high-profile social media platform at the moment, the JDD Lab is positioned to pivot and include any future social media platform that the public uses to look for help with legal problems. It is even possible, at least in theory, to accommodate aggregating other forms of data that use quite different methodologies.

Finally, with any assessment of something new, it bears remembering where Canadian legal scholarship and data analysis sit in relation to research in other disciplines and jurisdictions.

While Reddit data is a relatively unexplored area of research in law<sup>154</sup>, other fields – particularly medicine – have been collecting and analyzing Reddit data and other social media data for some time now. Mummun de Choudhury generated a set of “pioneering” papers<sup>155</sup> that used data from Twitter (2013)<sup>156</sup>, Facebook (2014)<sup>157</sup>, and Reddit (2014)<sup>158</sup> to monitor population health and

---

<sup>154</sup> In a 2021 study, 727 manuscripts were found that used Reddit data between 2010 and 2020, and of them only 2 were in law. Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler and Michael Zimmer, “Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics” (2021) 7:2 *Social Media + Society* 1.

<sup>155</sup> Mike Conway and Daniel O’Connor, “Social Media, Big Data, and Mental Health: Current Advances and Ethical Implications” (2016) 9 *Current Opinion in Psychology* 77 at 78.

<sup>156</sup> De Choudry et al used Twitter to ground truth their survey data. M De Choudhury, S Counts, and E Horvitz, “Social Media as a Measurement Tool of Depression in Populations” (2013) in *Web Science ’13: Proceedings of the 5<sup>th</sup> Annual ACM Webs Science Conference* 47-56, online: *ACM Digital Library* <<https://dl.acm.org/doi/proceedings/10.1145/2464464>> [<https://perma.cc/6JRB-YHU3>].

<sup>157</sup> M De Choudhury, S Counts, E Horvitz and A Hoff, “Characterizing and Predicting Postpartum Depression from Shared Facebook Data” (2014) in *CSCW’14 Proceedings of the 17th ACM conference on Computer supported cooperative work & social 626-638* 638, online: *E Horvitz* <<http://erichorvitz.com/FB-cscw2014.pdf>> [<https://perma.cc/9WGP-LHXV>].

<sup>158</sup> M De Choudhury and Sushovan De, “Mental Health Discourse on Reddit: Self-Disclosure, Social Support, and Anonymity” (2014) *Proceedings of the Eighth International AAI Conference on Weblogs and Social Media, Association for the Advancement of Artificial Intelligence* 8, online (pdf): *Proceedings of the International AAI Conference on Web and social media* <<https://cdn.aaai.org/ojs/14526/14526-28-18044-1-2-20201228.pdf>> [<https://perma.cc/2F2Y-8TEC>].

risk factors. Medical research has since considered how much and how well social media posts provide access to people’s first-hand experiences<sup>159</sup>, and have used Reddit data to track the public’s interest in and understanding of various health issues.<sup>160</sup> Researchers have also mined Reddit data extensively using both quantitative and qualitative analyses to gain information on topics ranging from anxiety during the recent pandemic, to common parenting issues<sup>161</sup>, to trends in skincare<sup>162</sup>.

Reddit data is also the focus of considerable work in the United States. In collaboration with the Suffolk Legal Innovation and Technology (LIT) Lab, Margaret Hagan and the team at Stanford University’s design lab used Reddit to update the country’s taxonomy of legal issues in 2018.<sup>163</sup> Hagan described the new list of 20 top-level problem types as more user-centred since it now includes the legal issues people post and say they are experiencing.<sup>164</sup> The group then developed a game called “Learned Hands” that invites people to spot legal issues in Reddit posts. At the same time, the people’s “game-play” supports Access to Justice artificial intelligence work.<sup>165</sup> Such is the value of the Reddit data that the team at Suffolk LIT Lab has used the data from

---

<sup>159</sup> N Alvaro, Conway M, Doan S, Lofi C, Overington J, and Collier N, “Crowdsourcing Twitter Annotations to Identify First-Hand Experiences of Prescription Drug Use” (2015) 58 J Biomed Inform 280.

<sup>160</sup> Andrew Park and Mike Conway, “Tracking Health Related Discussions on Reddit for Public Health Applications” (2017) AMIA Annu Symp Proc. 2018, 1362 (“Park and Conway 2017”).

<sup>161</sup> EM Westrupp, CJ Greenwood, M Fuller-Tyszkiewicz, TS Berkowitz, L Hagg, G Youssef “Text Mining of Reddit Posts: Using Latent Dirichlet Allocation to Identify Common Parenting Issues” (2022) 17(2) PLoS ONE e0262529, online: *PLOS ONE* <<https://doi.org/10.1371/journal.pone.0262529>> [<https://perma.cc/LG37-VL3C>].

<sup>162</sup> Jack A Cummings and Vinod E Nambudiri, “Natural Language Processing: A Window to Understanding Skincare Trends” (2022) 160 International Journal of Medical Informatics 104705, online: *Science Direct* <<https://doi.org/10.1016/j.ijmedinf.2022.104705>> [<https://perma.cc/BU73-L65R>].

<sup>163</sup> Hagan has been doing this work with David Colarusso from the Suffolk Lit Lab. Margaret Hagan, “Every Legal Problem That Exists: The Legal Help Taxonomy For Machine Learning” (September 4, 2018), online: *Medium* <<https://medium.com/legal-design-and-innovation/every-legal-problem-that-exists-the-legal-help-taxonomy-for-machine-learning-3c9f2f75460>> [<https://perma.cc/LP7F-S38L>] See also: LIST, “LIST = Legal Issues Taxonomy” online: *Taxonomy* <<https://taxonomy.legal>> [<https://perma.cc/9BAQ-T8NQ>].

<sup>164</sup> Margaret Hagan “A Human-Centered Taxonomy of Legal Problems” (June 5, 2018), online: *Medium* <<https://medium.com/legal-design-and-innovation/a-human-centered-taxonomy-of-legal-problems-fab415ebfe88>> [<https://perma.cc/CQ88-GQ9K>].

<sup>165</sup> Learned Hands, online: *Stanford* <<https://learnedhands.law.stanford.edu>> [<https://perma.cc/4V6C-A4UY>].

Learned Hands to build a computerized issue-spotting program, “SPOT”.<sup>166</sup> At the same time as the JDD Lab was pivoting to Reddit, Margaret Hagan and the team at Stanford University’s design lab were remarking favourably on the utility of Reddit as a forum to help users understand their rights and build their legal capability.<sup>167</sup>

Of course, Reddit has its own limitations, the greatest of which is that Reddit gathers no demographic information about Redditors. On Reddit, user profiles are optional. Users are only required to give a username, password, and email address. They are not required to share information that could identify them, such as their names, emails or locations (other than their home province or state).

There are some surveys regarding the general population of Reddit. They show that Reddit users tend to be younger and male. Ryerson’s Social Media Lab found that of Canadian adults online, more men than women reported having a Reddit account (18% of men to 13% of women).<sup>168</sup> The Social Media Lab also found the groups aged 18-24 years old and 25-31 years old have high percentages of Reddit use (41% and 30% respectively). Adults aged 35-44 years old were not far behind at 18%. However, for adults aged 45-55 and 55+, Reddit use dropped significantly (to 9% and 3%, respectively). Statistics from the USA report a similar picture: 23% of men and 12% of

---

<sup>166</sup> “The Legal Innovation & Technology Lab’s Spot API” (SPOT Version 2022 05 21 (Build 10), online: <https://spot.suffolklitlab.org>) [<https://perma.cc/3J62-363>].

<sup>167</sup> Margaret Hagan, “Legal Help on the Internet - Does Tech Live up to its A2J promise?” (April 24, 2021), online: *Access to Justice – Scholars Showcase, American Bar Foundation*, <<https://www.americanbarfoundation.org/events/19946>> [<https://perma.cc/Q5FA-3VER>].

<sup>168</sup> Philip Mai and Anatoliy Gruzd, “The State of Social Media in Canada 2022: A Census-balanced Survey About Social Media Adoption and Use in Canada” (Toronto: Social Media Lab Toronto Metropolitan University, 2022) at 16.

women polled reported using Reddit<sup>169</sup>, while 36% of people using Reddit are between the ages of 18-29 and 22% are between 30-49. Again, as age increased, Reddit use declined: in people aged 50 -64, only 10% reported using Reddit and that number shrank to 3% for those over 65 years of age.<sup>170</sup>

However, research shows subreddits are distinct from one another. Each subreddit creates its own community of Redditors who wish to engage a topic, and topics vary widely. A 2019 study reviewed 100 subreddits covering topics including general interest, humour, news/politics/religion and gaming. It found the percentage of women in the subreddits ranged from 12.6% to 65.0%.<sup>171</sup> Within Reddit, there have been some surveys of certain subreddits, but not of the subreddit “r/legaladvicecanada”.<sup>172</sup> Thus, while we should take note that the general population on Reddit is male and younger, we cannot say definitively what population of people form the community on “r/legaladvicecanada”.

While these numbers offer some description of who is active and posting to Reddit, we must also consider who we are not seeing in the data. The general demographics of Reddit suggest we will see few posts from older people with legal problems. Anecdotal evidence in BC suggests that

---

<sup>169</sup> Statista, “Percentage of U.S. adults who use Reddit as of February 2021, by gender” (2021), online: *Statista* <<https://www.statista.com/statistics/261765/share-of-us-internet-users-who-use-Reddit-by-gender/>> [<https://perma.cc/Z2GS-3L4N>].

<sup>170</sup> Statista, “Percentage of U.S. adults who use Reddit as of February 2021” (2021), online: *Statista* <<https://www.statista.com/statistics/261766/share-of-us-internet-users-who-use-Reddit-by-age-group/>> [<https://perma.cc/77GB-ZA6T>].

<sup>171</sup> Mike Thelwall and Emma Stuart, “She’s Reddit: A Source of Statistically Significant Gendered Interest Information?” (2019) 56:4 *Information Processing and Management* 1543 at 1553.

<sup>172</sup> The subreddit “r/Canada” surveyed its membership in 2019 prior to the general election held in October of that year. They found an overwhelming percentage of their Redditors identified as male (84% male to 10% women) and between 25-39 years of age (62%). “r/Canada”, online: <[https://www.Reddit.com/r/canada/wiki/2019survey/#wiki\\_subreddit\\_demographics](https://www.Reddit.com/r/canada/wiki/2019survey/#wiki_subreddit_demographics)> [<https://perma.cc/B2A9-C7LB>].

Indigenous populations, particularly those who live in remote communities are more likely to use a Facebook group, or even a radio channel, to connect virtually as a group and ask for support with problems.<sup>173</sup> I offer these examples not as a comprehensive assessment of who might be excluded from this dataset, but rather as a marker to ensure we take into account whose everyday problems are less likely to be represented in this data.

There is no denying there would be value in having demographic data regarding the Redditors posting and looking for legal advice. Such data could open up entire avenues of research regarding, for example, what parts of the population are looking for what kinds of help. However, medical researchers working with Reddit data have remarked that the anonymity afforded by Reddit may have a positive effect on data quality. When tracking health-related discussions, researchers suggested the anonymity means Redditors are more likely to share specific details in their comments.<sup>174</sup> For example, the anonymous aspect was commented on favourably when Reddit data was used successfully to track health related discussions on Reddit about controversial topics and identify opportunities to improve health education and thus potentially, health outcomes<sup>175</sup> and when Reddit data analysis identified patterns during the Covid-19 pandemic.<sup>176</sup> I also note Genn's early comment that certain types of problems generate different types of efforts towards resolution. For example, for certain family matters people

---

<sup>173</sup> Personal communications with Patricia Barkaskas, Strategic Advisor to the Dean for the National Centre for Indigenous Laws, Faculty of Law, University of Victoria (February 14, 2023).

<sup>174</sup> Park and Conway 2017, *supra* note 160 at 1363.

<sup>175</sup> Park and Conway 2017, *supra* note 160.

<sup>176</sup> Low et al. 2020, *supra* note 151; Allie Slemon, Corey McAuliffe, Trevor Goodyear, Liza McGuinness, Elizabeth Shaffer and Emily K. Jenkins, "Reddit Users' Experiences of Suicidal Thoughts During the Covid-19 Pandemic: A Qualitative Analysis of r/Covid19\_support Posts" (2021) 9:693153 *Front. Public Health* 1 ("Slemon 2021"); Corey McAuliffe, Allie Slemon, Trevor Goodyear, Liza McGuinness, Elizabeth Shaffer, Emily K. Jenkins, "Connectedness in the time of Covid-19: Reddit as a source of support for coping with suicidal thinking" (2022) 2:100062 *SSM Qual Res Health* 1.

engage with the formal justice system, while for consumer problems people do not. Reddit data contains valuable information on whether the help requested and required was part of the formal justice system or not, something Sandefur has already explained has great value.

Another limitation to keep front of mind is the self-selection bias inherent in social media data.<sup>177</sup> The most recent survey shows that most people with an everyday legal problem take some action to resolve the problem, and that the top place they go looking for help is the internet. There are likely many places where people go on the internet, but we know a top place is Reddit (at least in BC). We cannot say it is a representative sample of the average Canadian or British Columbian's everyday legal experiences. That kind of work requires carefully planned survey methodologies, and significant time and funding, and is underway elsewhere. However, one legal scholar has already pointed out that data from Reddit can and should be used just as any other community research data would be used.<sup>178</sup> We can say, then, that the growing Reddit dataset at the JDD Lab illuminates the problems of people from British Columbia who use Reddit. We can also manage some of this bias by gathering posts concerning justiciable events from subreddits that are titled after cities and regions, thus gathering a larger set of people who choose to post to Reddit for a variety of reasons.

---

<sup>177</sup> Moore, Lisa, "Measuring Impacts of Legal Services – A Literature Review on Research Design and Methodology" (2020) Canadian Forum on Civil Justice 9 ("Moore 2020") at 10, online: *Osgoode Digital Commons* <<https://digitalcommons.osgoode.yorku.ca/cgi/viewcontent.cgi?article=1009&context=cfcj>> [<https://perma.cc/4R4X-PKYD>].

<sup>178</sup> Dylag, Matthew, "Measuring Access to Civil Justice: An Empirical Study of Ontario's Reform Initiatives"(2021) PhD Dissertations 75 ("Dylag 2021") at 86, online: *Osgoode Digital Commons* <<https://digitalcommons.osgoode.yorku.ca/cgi/viewcontent.cgi?article=1075&context=phd>> [<https://perma.cc/5AHD-KQ9Y>].

After assessing the strengths and weaknesses of Reddit data and those of Everyday Legal Needs survey data, Reddit data comes out as a valuable data source. Its limitations can be managed so that the justice system may benefit from the enormous value of this data for providing researchers person-centred, evidence-based information on how to improve access to justice. Everyday Legal Needs data may be the “dominant” form of data for assessing people’s problems and what they do about them, but if the justice system is serious about taking a person-centred approach to improving access to justice, then Reddit data presents a complimentary and desirable data source.

## CHAPTER 5: CONCLUSION AND FUTURE WORK

### Conclusions

To solve Canada's access to justice problem, we need data about what problems people face, what they do to try and resolve them and what help they need. In this thesis, I have reviewed the dominant data form that offers this information – Everyday Legal Needs surveys – and I have introduced new data from the social media site, Reddit. The person-centred data generated by Everyday Legal Needs surveys is invaluable evidence for those who work on justice policy and social policy generally. However, these surveys have significant weaknesses. I have used my thesis to show how Reddit data, subjected to the kinds of tools and analysis described above offers a complimentary data source. Reddit data does not experience the same weaknesses as survey data and its strengths are numerous and still unfolding.

This thesis offers some lessons to the Access to Justice community for how to navigate this information space. First, when using Everyday Legal Needs survey data, researchers must be keenly aware of the methodological assumptions built into the data and any changes in methodology in a survey series. It will be interesting to see if Statistics Canada's next survey maintains the methodology adopted in its 2021 survey of Canada's Legal Problems, because that would suggest the beginning of a consistent dataset which researchers could use to spot trends in Canadian's access to justice. LABC has two comparable survey datasets from its surveys in 2018 and 2020 and hopefully it will generate more soon.

Second, principal takeaways from this thesis regarding Reddit are:

1. Reddit data is generated by people who chose, in their moment of need, to share what they need on a social media platform and ask for help. It is the epitome of person-centred data. Even better, Reddit data is available and ready to be used to support evidence-based action that improves access to justice.
2. Reddit offers new and unexpected information on how people's everyday legal problems cluster together. Since problem types in the Reddit data "bubble up" organically within the JDD Lab's collection and analysis processes, Reddit data shows us how people's problems cluster organically, without being gathered into pre-determined categories. Further, Reddit data can provide information on problem types that might improve future surveys.
3. The immediacy of Reddit data complements survey data by filling in the data gap between surveys. Further, the regular analysis of Reddit data can reveal problems in realtime, as they "bubble up" in people's lives. In turn, this can generate opportunities to better support users who are experiencing a particular, timely challenge.
4. Reddit data is being used by others, particularly those in other disciplines, including medicine. It is also being used by legal scholars in the USA.

When Sandefur discussed research on what help people are asking for with their legal problems, she concluded her paper by calling for new approaches on the access to justice problem from different disciplines and problem-solvers other than those already embedded in the legal system.<sup>179</sup> She emphasized the value of a robust evidence base, and encouraged using it to learn more about when formal justice processes are necessary and when they are not. The JDD Lab's

---

<sup>179</sup> Sandefur 2019, *supra* note 27 at 54.

work with Reddit is a testament to the validity of her call to action. The Lab’s experimental nature created a space for a collaboration between data scientists and legal scholars, and then drove them – by necessity – to develop a new source of access to justice data, to identify and apply new analysis to that data and, with this thesis, to consider the value of that data. This is important, innovative and necessary work on the access to justice problem. It has brought forward data that opens a new “window” onto what help people are asking for with their legal problems.

### Future work

Building on lessons identified in this thesis, the JDD Lab is continuing its work:

1. The JDD Lab continues to grow the Reddit dataset, collecting approximately 300 new posts each month. The Lab plans to share monthly data updates with the access to justice community which will include interactive LDA and intertopic distance maps for recipients to review. Given the collection and aggregation structure outlined in Chapter 3, it is relatively simple to write further code for the research engine, which would collect Reddit data for each of Canada’s provinces, not just British Columbia. As the Lab builds relationships with other groups from other provinces, it is in a position to provide that data to other teams. We hope these collaborations will generate further insights and ideas for research as more people turn their attention to analyzing Reddit data, something we believe will benefit all involved.
2. As the Reddit dataset grows, the JDD Lab will continue to look for further examples of issues “bubbling up” in the Reddit data which could signal opportunities to provide better support and resources to users experiencing a particular, timely challenge.

3. The JDD Lab will conduct further analysis on the clusters which LDA is already creating in an attempt to identify “subclusters”. The goal is two-fold: first, and in keeping with the JDD Lab’s goals, to identify issues within a cluster that could signal an opportunity to improve access to justice. Second, to learn more about how the data complements Everyday Legal Needs survey data. For example, if we work with a cluster which matches the problem types from surveys, how will the subclusters within it relate to the “prompting questions” used in survey methodology? We believe this could be valuable information for future surveys undertaken in British Columbia and Canada.
4. The JDD Lab will subject the data to a supervised clustering method - a Support Vector Machine (SVM) – to allow comparison between these supervised clustering results and the unsupervised clustering results from LDA and Inter-topic distance mapping. This will involve using the prompting questions as “sub-problems” in order to train the SVM to spot posts which should be allocated to a “sub-problem”. Our research objective would be to learn more about the value of both clustering methods. One hypothesis might be that the results from the SVM clustering and sub-clustering will build a dataset of information on the problem types and sub-problem types used in surveys. This would mean we could look for trends month to month or even on certain days of the week. We would also be in a position to assess if the data showed any reaction after changes are made to elements of the justice system.
5. To address some of the limitations discussed in this thesis, we will use the Reddit data itself to provide some demographic information about Redditors. For example, using a simple list analysis, it may be possible to learn more about the location from which people are posting, be it urban or rural. We also plan to collect Reddit data from

subreddits other than “r/legaladvicecanada”. An example would be subreddits that describe their Reddit community with locations, such as “r/vancouver”. The JDD Lab is enthusiastic about developing coding to identify posts which contain a justiciable event within location-based subreddits.

6. The JDD Lab is not limited to Reddit data, and I will be looking into other data sources, particularly those which can provide information on groups we may not be seeing in the Reddit data. The JDD Lab is particularly interested in gathering data from new sources such as Facebook or radio channels which could be used to identify opportunities to improve access to justice for remote Indigenous populations.
7. Finally, the JDD Lab’s success at securing and analyzing Reddit data puts the Lab in a good position to generate opportunities where new tools could provide people with the help they are asking for. For example, the evidence presented in this thesis shows that most people in Canada looking online for legal advice are going to Reddit. The evidence-based, person-centred opportunity to improve people’s access to justice would be to meet people where they are going and offer them the help they are asking for. In this case, that would mean creating a prototype which could meet people on Reddit, understand what help people are asking for, and provide that help. The prototype would use Artificial Intelligence and Large Language Models, such as ChatGPT, to generate the help. Careful attention would be paid to build a prototype that uses AI with “guard rails” to ensure accuracy and transparency in the help provided. This project is already underway, and I look forward to reporting on this work soon.

## Appendix 1

**Table 1: Problem Types in 6 Everyday Legal Needs Surveys**

Note: Problem types are presented in the order they appear in each survey. Where possible, problem types have been edited to a standard heading for that problem type, to support cross-referencing between surveys.

	1996	2004	2006	2008	2013/2014	2021
	Genn <sup>180</sup>	Currie <sup>181</sup> (Low- and Moderate-Income Only)	Currie <sup>182</sup>	Currie <sup>183</sup>	CFCJ <sup>184</sup>	Statistics Canada <sup>185</sup>
	In-person	Telephone	Telephone	Telephone	Telephone	Online
1	Employment	Consumer	Consumer	Consumer	Consumer	Consumer
2	Owning residential property	Money and debt	Money and debt	Money and debt	Money and debt	Employment
3	Renting rooms or property	Employment	Employment	Employment	Employment	Personal injury/serious health issue
4	Living in rented property	Social/income assistance	Social/income assistance	Social/income assistance	Social/income assistance	Neighbourhood
5	Consumer	Disability benefits	Disability benefits	Disability benefits	Disability benefits	“Your house, your rent, your mortgage or rent owed to you”
6	Money	Housing	Housing	Housing	Housing	“Money you owe (debt) or money owed to you, excluding rent and housing issues”
7	Relationships, other family matters	Immigration	Immigration	Immigration	Immigration	“Getting social or housing assistance, receiving Old Age Security, Guaranteed Income Supplement or other government assistance

<sup>180</sup> Genn 1999, *supra* note 7.

<sup>181</sup> Currie 2006, *supra* note 59; Ab Currie, “A National Survey of The Civil Justice Problems of Low and Moderate Income Canadians: Incidence and Patterns” (2005) Department of Justice: Canada, online (pdf), *CFCJ*: <<http://www.cfcj-fcjc.org/sites/default/files/docs/2006/currie-en.pdf>> [<https://perma.cc/YJK2-ZUMY>].

<sup>182</sup> Currie 2007, *supra* note 36 at 6.

<sup>183</sup> Currie 2009a, *supra* note 60.

<sup>184</sup> Farrow et al 2016, *supra* note 84.

<sup>185</sup> The Canadian Legal Problems Survey does not use headings for all problem types. This table uses “standard headings” where possible and in all other cases uses language from the “Problem identification” section of the survey. StatsCan Survey, *supra* note 95.

						payments, or problems with the amount received”
8	Children	Discrimination	Discrimination	Discrimination	Discrimination	Disability assistance
9	Injuries, health from accidents	Treatment by Police	Treatment by Police	Treatment by Police	Treatment by Police	Immigration
10	Discrimination	Family problems (relationship breakdown – breakup, divorce, child custody)	Family problems (relationship breakdown – breakup, divorce, child custody)	Family problems (relationship breakdown – breakup, divorce, child custody)	Family problems (relationship breakdown – breakup, divorce, child custody)	Contact with police
11	Unfair police treatment	Other family	Other family	Other family	Other family	Contact with police as a victim or witness, excluding neighbourhood problems.
12	Immigration	Wills and power of attorney	Wills and power of attorney	Wills and power of attorney	Wills and power of attorney	Family breakdown, excluding child custody
13	Negligent Medical/dental treatment	Personal injury	Personal injury	Personal injury	Personal injury	Child custody
14		Hospital treatment or release	Hospital treatment or release	Hospital treatment or release	Hospital treatment or release	“A will, or taking care of financial or health issues for a person who was unable to look after Themselves”
15		Threat of legal action	Threat of legal action	Threat of legal action	Threat of legal action	Poor or incorrect medical treatment
16			Neighborhood problems	Neighborhood problems	Neighborhood problems	Threat of legal action
17					Criminal charges	Harassment
18						Discrimination