JUDGMENTS AS DATA
AUTOMATED OPEN-ACCESS ANALYTICS
FOR DECISIONS OF COURTS AND
TRIBUNALS IN NEW ZEALAND

December 2020
Tom Barraclough | Curtis Barnes | Warren Forster
# Overview

Open access to digital case law in New Zealand

- Previous access to justice research (2015)
- Partnership with OpenLaw NZ (2019)
- Access to digital case law in New Zealand
- Summary of our findings

## Part One

Accessing judgments at scale

- Why we should enhance the availability of judgments
  - Judgments state the law
  - Constitutional factors, Open Justice and Freedom of Expression

Acknowledged limitations on public access to judgments at scale

- Privacy and sensitive cases
- Maintaining judicial discretion and control (including suppression)
- Would another system better protect these principles?

How is case law currently accessed outside of commercial publishers?

- PDF and “access” as machine readability
  - Distribution of case law as PDF by email
  - The New Zealand Council of Law Reporting

Conclusion to Part One

## Part Two

Balancing benefits and trade-offs

- The benefits and applications of AOAA and judgments as data
  - Effective public databases of case law
  - Linked legal materials
  - Citators
  - Tracking suppression and identifying information subject to suppression
  - Research platforms for empirical work
  - Legal education opportunities
  - Innovation in legal technology enterprises

Predictive analytics and machine learning on case law

- Predictive analytics on case law
- Analysing the behaviour of participants in justice system
- Why these potential uses should not prevent development of the system
APPENDIX FIVE: JUDGMENT DISTRIBUTION SHEET .................................................................................. 101
APPENDIX SIX: DISCUSSION PAPER ON “JUDGMENTS AS DATA” .................................................. 106
PREAMBLE

This report addresses access to case law in New Zealand, why it is important, and how it can be improved, particularly for use in digital services. The report should be read against the following context:

- We acknowledge there are competing interests and priorities for resources in the justice system. Our recommendations here should not be taken as trumping all other interests.

- The justice system and people working in that system have important everyday responsibilities which cannot be compromised. The importance of these responsibilities must be weighed against the recommendations and analysis in this report.

- We emphasise that providing improved access to primary legal materials is only one factor within the overall outcome of increasing access to justice, alongside other features such as enhanced access to legal advice and to the Courts.

- Nothing in this report is intended as a criticism of the judiciary or of its individual members.

ACKNOWLEDGEMENTS

This report and the OpenLaw NZ software proof of concept were generously funded by Te Manatū a Ture o Aotearoa, the New Zealand Law Foundation. We offer our thanks to the Law Foundation Board and to NZLF Executive Director Lynda Hagen. We also extend our thanks to the NZLF Information Law and Policy Project Advisory Review Committee and the project manager of ILAPP, Richman Wee.

Our thanks to the Accident Compensation Corporation for sharing copies of judgments used in developing the OpenLaw NZ proof of concept. We also wish to acknowledge the valuable work done by the New Zealand Legal Information Institute (NZLII), the Australasian Legal Information Institute (AustLII) and the wider Free Access to Law Movement over a long period.

Thank you to OpenLaw NZ and its directors and co-founders Andrew Easterbrook and William Parry. Their determination to improve access to justice in New Zealand is both inspiring and greatly appreciated, as are the many long days and nights they have committed to this objective. Thank you also to the volunteers for OpenLaw NZ who participated in development of the tool: their contributions are recorded on OpenLaw NZ website and on the OpenLaw NZ GitHub repository.

We also thank the various people who have engaged with us or provided support in the course of producing this report. They have not seen the final report and any errors are attributable only to the authors.
EXECUTIVE SUMMARY

1. In this project we considered digital tools for accessing, analysing, and distributing case law (understood as written decisions of courts and tribunals in New Zealand). We focused on public access to such tools and to case law generally. We refer to these digital uses of case law generally as “automated open access analytics” (AOAA).

2. Case law is important because it is one of two primary sources for learning about and understanding what the law of New Zealand is and what it requires people to do. It is also the method by which judges and tribunals explain the reasons justifying their decisions about how the law applies in an individual case. In this way, it also contributes to judges’ accountability in the judicial system and to the public.

3. Advanced automated open access analytics have many potential uses. These include building basic access mechanisms that let people read cases linked to each other in a simple way, as well as more sophisticated applications of algorithmic processing and machine learning. They treat written decisions of courts and tribunals as digital data, which can be processed using computer systems for a wide range of purposes.

4. We asked whether an automated open access analytics platform (AOAA) that uses judgments as data could be built. We concluded that it can. OpenLaw NZ is one such platform among others that could exist.

5. While AOAA systems like OpenLaw NZ can be built, building the technology platform is only one part of the overall process. Anyone who wants to build AOAA will also need access to judgments in large volumes, “at scale”, in a manner that facilitates digital processing.

6. There are both principled and practical reasons why it is difficult to publish judgments generally. For instance, publishing carries risks for the publisher and others. Case law is not like statute law because it includes information about real people in dynamic situations. Editorial responsibility must be exercised when judgments are published, whether at the point of creation or publication. Historically, this has led to delegation of this editorial judgment to commercial publishers. As a result, users of proprietary commercial systems currently have superior access to case law than do other members of the public. They also access that case law using systems that enhance their ability to engage with it.

7. Subsequently, the way that judgments are published by the courts and by publishers reflects an attempt to navigate these challenges pragmatically, but not optimally. The result is that access is inadequate not just at scale, but also for everyday users accessing one decision at a time. Issues include access to databases, completeness of those databases, functionality of those databases, and machine readability of judgments, including their quality as a data source for use in computer systems. As a result, attempts to build AOAA in New Zealand will face significant difficulties if current arrangements persist.
8. We acknowledge that some applications of AOAA to case law may pose a risk of harm to individuals or the administration of justice if used recklessly, however we say these risks can be managed. Further, we say that these kinds of risks already exist under the current system, and can be better managed by using more effective digital systems.

9. We describe the features of a potential system for preparing, distributing and analysing case law for digital systems in this report. We refer to this as "a system", but it is really composed of organisational, operational and digital processes.

10. What we suggest is that case law be produced from the beginning with a view to its use in AOAA systems. We suggest the adoption of processes that prepare judgments using standardised metadata, using tools like extensible markup languages (XML) and standards like Akoma Ntoso (AKN). We point to ways this can be done with minimal disruption to the current system and to judicial workload.

11. Adoption of our proposed system would enable enhanced access to case law for the public on a level that is currently only available to people with commercial subscriptions to legal publishing companies. Our system would also enable the judiciary to exercise enhanced control over judgments using access-controlled systems. This has benefits for contemporary issues in the justice system, such as how to track the duration and scope of suppression and non-publication orders.

12. New Zealand should take a coherent and systematic approach to making the written decisions of courts and tribunals available to the public free of charge. The current system is unjustifiably splintered between multiple government sources, NZLII, and commercial publishers. In particular, New Zealand should ask whether it is acceptable that its official law reports are only available in exchange for monetary consideration through a commercial publisher. There is public interest in having the systems by which judgments and written decisions are produced and distributed mapped and made available for public scrutiny and policy analysis.

13. In light of deficiencies in current access to case law in New Zealand, contemporary issues such as digital publishing and name suppression, and the benefits that could result from a reformed system, we say New Zealand should adopt our suggested approach. We say that this would simply be an evolution of existing judicial function and practice, not a revolution, with wide benefits for the public and the judiciary.
A NOTE ON “STRUCTURED DATA”

14. In this report we advocate for a system where decisions of courts and tribunals in New Zealand are created, stored and published as “structured data”. At the outset, we wish to explain what this means in simple terms.

15. Currently, judgments and written decisions are prepared predominantly as text for human consumption. To a person observing a page of text, the text is structured using visual cues. There are clear signifiers that distinguish between the kinds of text or data stored on the page. There is also consistent visual formatting that places, for example, the names of the parties, the date of the judgment and the name of the judge spatially on the front page. Text is further organised into paragraphs, which are numbered.

16. At the same time, that text is capable of being processed by computers. In the process of extracting text from PDF format (the standard in publishing written decisions), much of the structure is lost. Further, most of what confers structure on the document for a human reader relies on human units of meaning that, primarily, are observed visually. To a machine, however, much of that structure is not useful.

17. This problem could be resolved by the addition of other structural elements in the documents that comprise written judgments. These new structural elements are digital: they can be read and processed by computer systems. This would make them amenable to more detailed computational analysis, as well as different kinds of human analysis.

18. What we are really suggesting is the use of metadata (data about data) that tells a computer system and the humans using that system how to understand the significance of the text available in a decision. This added functionality does not require any change to the visual appearance or presentation of case law to a human user.

19. At present, the OpenLaw NZ platform and others like it, such as NZLII and AustLII, use automated systems to infer that a particular string of text appears to be a citation, or a name, or a date. This is primarily based on its syntax, or the formulaic expression of those textual fields (e.g. the relatively predictable way in which a date is expressed in written language). With written language being what it is, there may always be rare exceptions where something that “looks like” a citation or a date (for example) does not fit the parameters of the computer as programmed. That leads to what is referred to as a “false negative”: something that should have been captured was missed. For instance, a date may be expressed in a manner not anticipated by the programmers or a citation may be incorrectly formatted. The inverse of this is a false positive: that is a situation where an algorithm detects a string of text that “looks like” a particular kind of data (a name, a date or a citation), but is actually not what the programmer intended to identify. This creates a risk of error that leads human users to perceive such automated systems as unreliable.
20. Our essential conclusion is that these problems can be avoided by preparing the judgment in a way that anticipates subsequent digital processing. As such, judgments would be produced as structured machine-readable data, as well as the unstructured data that is familiar to humans, but more opaque to computers.

21. New Zealand has already commenced along this path of structural formalisation in legal documents. For example, the New Zealand Law Style guide is essentially a standard to govern the structuring of textual information so that it will be consistent, easier to process, and less likely to be misinterpreted. Further, we say that the judiciary is already actively engaged to some degree in the editorial curation of individual decisions and published law reports in New Zealand, as is the Ministry of Justice. Accordingly, the changes we suggest are not a revolution, but an evolution of current practices.

22. To everyday users of the justice system, including the judiciary, this significant change in the functionality of written decisions and the way they are prepared may not lead them to look any different than they do now. But as a result, free public access to a wider range of judicial decisions in more useful formats will be massively increased. Further, through the intelligent use of digital systems, the judiciary’s ability to exercise lawful influence over its primary output will be amplified, according to the interests of justice.

23. This presents an exciting opportunity for the New Zealand justice system to exercise cautious and judicious leadership. We have discovered that other common law systems are just beginning to turn in this same direction. These technological methods are well established, even if their application in this context may be novel. Importantly, the impetus behind this development is grounded in longstanding legal principles like the rule of law and the importance of access to justice. Equally, when judgments are available as structured data, this will support a range of innovations to resolve contemporary issues facing the justice system.

24. In any event, the current state of inequitable access to written decisions as primary legal materials cannot reasonably be allowed to continue. Preparation of judgments in this way is the most sensible method of durably remedying this inequity.
OVERVIEW

OPEN ACCESS TO DIGITAL CASE LAW IN NEW ZEALAND

PREVIOUS ACCESS TO JUSTICE RESEARCH (2015)

25. The following description provides context for how we arrived at this topic and our subsequent conclusions.

26. In 2015 we conducted legal research into the barriers people face when disputing decisions of the Accident Compensation Corporation through the courts.¹ To do that, we used a team of five people to read, analyse, and collect information from more than 500 judicial decisions.² We adopted a mixture of qualitative and quantitative methods. Our focus was on using the judgments themselves as a data source. Despite extensive legal training, the research task was manually intensive, time consuming, and required a sound grasp of subject matter and technique to conduct.

27. That research was instrumental in shaping public policy.³ We think similar methods can be used by government officials, academics and independent researchers to learn important things about other areas of our legal system, with important insights from a regulatory stewardship perspective.⁴

28. The 2015 research reflected our contemporaneous opinion that there were significant access to justice problems in the ACC dispute resolution process. The same was true for the policy system that fed that dispute

¹ Acclaim Otago Inc "Understanding the problem: An analysis of ACC appeals processes to identify barriers to access to justice for injured New Zealanders" (9 July 2015).
² In this report we refer to “case law”, “written decisions”, and “judgments” interchangeably. At times, we also include written orders of the Court of general public effect, such as suppression or non-publication orders. While there are differences between these things, they have in common the quality of being written records of the factual findings, reasons, and legal analysis followed by a judge or a statutorily empowered decision-maker like a tribunal in reaching a decision.
resolution process. Our opinion was informed by litigation experience within the ACC jurisdiction, and by talking with those impacted by the systemic problems. We noted that there was little extant empirical data that could corroborate our opinion at that time.

29. We subsequently identified judicial decisions as a publicly available source of information which might provide a sound, factually reliable record of individual cases. We hypothesised that by processing these individual judgments on a large scale, and by recording pertinent information in a methodological way, we would be able to generate the data we needed to prove the existence of access to justice barriers across the system. As such, we used judgments to produce useful social data that would prove our hypothesis. This was our first meaningful exposure to the value that open access to legal data could have for improving access to justice at scale.

**Partnership with OpenLaw NZ (2019)**

30. In 2019, members of the same team partnered with OpenLaw NZ with funding from the New Zealand Law Foundation. OpenLaw NZ is a charitable organisation that has built an open source software platform. “Open source” means that the platform’s computer code is all publicly available to be scrutinised by anyone who wishes to do so, and can also be freely adopted within some permissive licencing conditions.\(^5\)

31. The OpenLaw NZ platform takes written judgments in PDF format, extracts the text from those PDFs, and uses computational processing (algorithms) to detect patterns in that text. What that means is that the OpenLaw NZ platform can use automated processes to detect important information from the text of thousands of judgments at a time, such as dates, citation information (references to legislation and case law), keywords and phrases, the name of the judge, the name of the parties, and other material information. The kind of information that can be extracted is limited only by the capacity of algorithms to detect it and the limitations of the source material (judgments). Once that text has been extracted and processed, it is available as structured data.

32. When judgments are processed in this way and made available as structured data, analytical and statistical processes can be applied to that data for a wide range of purposes. In this project, we refer to such capabilities as automated open access analytics, or “AOAA” for convenience.

33. Importantly, because OpenLaw NZ’s platform is open source, this means that anyone else can work independently or with OpenLaw NZ to create

their own algorithms and answer their own research questions using the
platform.

34. In this project, we have focused on the concept of “open access” because
we found that, in New Zealand, the only way to gain access to a
comprehensive collection of case law (and the digital means of
navigating it in useful ways) is to obtain a paid subscription to
commercial platforms. While other free resources exist, and in the case
of NZLII they are excellent, there is no resource for case law in the way
that Parliamentary Counsel provide in relation to legislation. None of
them permit researchers like us, or organisations like OpenLaw NZ, to
extract large volumes of case law all at once for machine analysis. This
is remarkable given that judgments of the courts are a public good which
also materially affect the lives of all New Zealand citizens subject to the
rule of law.

35. The kind of information that OpenLaw NZ’s platform can extract
replicates some of what we focused on in our previous manual research
published in 2015, particularly when it comes to automatically extracting
references to legal material.6 For example, being able to identify dates is
a basis for generating timeframes to infer how long disputes take to
reach resolution. Being able to identify the parties and their
representatives is a basis for inferring how many people are going to
court without a lawyer. We explain the process we adopted and some of
the findings we were able to reach in an appendix to this report.7

36. Subsequently, in this project, we asked ourselves what would be required
to work with OpenLaw NZ or organisations like it to create an automated
open access analytic tool for New Zealand case law that would help to
generate valuable insights about the legal system. Specifically, we
wanted to conduct similar research to the kind we did in 2015, except this
time, with the benefits of scale and speed provided by digital
technologies.

37. In that 2015 research, because of how many cases there were,
sometimes we had to rely on findings from a randomised sample of
cases. Using computational processing, it is feasible to process every
every judgment we are able to access: in this case, OpenLaw NZ’s platform can
process all 7000-odd cases provided to us by the Accident
Compensation Corporation and generate statistical data about them.

38. We hoped we could derive similar or better insights than we had done
previously, and hoped to learn new things about systemic legal research
in the process, however such insights were not our only goal. Our focus
expanded to consider what the barriers were to the overall process.

6 See the Understanding the Problem report and T Mijatov, W Forster, T
Baraclough “Problems with access to law in personal injury disputes”
7 Appendix Two.
ACCESS TO DIGITAL CASE LAW IN NEW ZEALAND

39. During the project, it became clear that developing superior AOAA systems was not possible without considering public access to judgments. These judgments are the source of data upon which AOAA relies. As such, our research began to consider how policy systems may encourage or inhibit access to case law as the principal catalyst for developing AOAA systems and making them available to the public free of charge.

40. In our attempt to understand how case law is accessed in New Zealand, we asked and answered the following questions:

- How do people without commercial subscriptions get access to case law in New Zealand?
- What is required to build an automated open access analytic platform for case law?
- What else can be done with that platform?
- What are the barriers to building it?
- How do people who want to analyse case law in large volumes get access to it?
- How are judgments produced and distributed in New Zealand and how might this be affecting public access to them?

41. From our work as legal researchers and as legal advisors to clients, we were aware of the vastly different level of access to case law offered through freely available databases compared to the level of access offered through large multinational commercial publishers. We commented on this at the time we were conducting our 2015 research. We also describe a specific situation where cases on a question of tenancy law could not be accessed using public databases.

42. In the present project, we emphasise that there are two kinds of “access” being contemplated.

a. The first relates to bare access to primary legal materials, including complete and accurate databases of all relevant case law someone might need to conduct legal research using decisions of various courts and tribunals.

b. The second kind of access relates to the usability of that case law in digital systems, including by access to the kind of digital tools that enable hyperlinking between primary legal materials or more generally facilitate their use in digital applications.

---

8 See Understanding the Problem at para 231.
9 See Appendix Four to this report.
In New Zealand, neither of these kinds of access are in an acceptable state.

Beyond “access” there is “accessibility” as it relates to disabled people. For people with visual impairments, PDF is generally an inaccessible format. What this means is that, without additional accommodations, disabled people are precluded from the first kind of access to an even greater degree than people without visual impairments. Further, disabled people may use digital assistive technologies to access case law. These assistive technologies rely on the second kind of access described above: usability in digital systems. The current state of access to case law therefore has a discriminatory effect on disabled people in New Zealand.

**SUMMARY OF OUR FINDINGS**

We came to an interesting (and perhaps surprising) conclusion, but one that we have found reflected in recent international academic commentary, new public policy initiatives in the United Kingdom, and civil society initiatives in Canada.

In summary: yes, it is possible to create an automated open access analytics platform for New Zealand case law. However, we found that there are significant barriers to creating and delivering such a system because of the way the system of publishing case law currently operates.

One of the major practical barriers to AOAA is the state of inconsistent and limited access to case law at all, let alone whether it can be accessed at scale using automated methods. We fully acknowledge the reasons that case law is not published as openly and fully as statute law, as a matter of law and public policy. These include editorial considerations that cannot be avoided and are a necessary part of the administration of justice, such as name suppression, and the privacy of the parties and witnesses. But these considerations are not an insurmountable barrier to providing access to case law in volumes and formats adequate for scaled computational analysis.

In fact, we suggest that those considerations could be better dealt with through a re-designed system of legal publishing, which we outline in this report. Our suggested system centres the role of the judiciary as ultimate

---

editors and supervisors of the way case law is published and used. This system has legal and constitutional benefits irrespective of whether it is used to implement AOAA.

49. The current state of affairs also creates an unacceptable inequity of access to primary legal materials in both bare access and usability. Put simply, people with commercial subscriptions to case law databases get access to written decisions of the courts that others do not. They also get access to vastly superior tools for navigating those written decisions. Where it is possible to remedy this using modern computational technology, we ought to do so. The current system is an anachronism from a period in which the legal system was shaped by paper documents, and deserves to be updated. A future system is possible utilising digital technologies and publishing processes that have developed since the advent of the modern internet in the late 1990s and the wide adoption of standardised mass communication technologies in the last decade.

50. In conclusion, we outline a specific set of digital and operational systems and processes that would provide improve access to case law, provide better tools for using and processing that case law, and also better protect the important interests people have in a well-functioning legal system, such as suppression, protection of privacy and the principle of open justice.
PART ONE

51. In this part we set out how case law is currently accessed by the public using digital systems. We also briefly cover the reasons why access to case law is important.

ACCESSING JUDGMENTS AT SCALE

52. Any platform intended to provide people with the ability to digitally analyse judgments in large volumes will need access to large quantities of judgments. In New Zealand, it is not possible to gain access to large quantities of judgments without seeking and obtaining the cooperation of a larger institution. For our project with OpenLaw NZ, we approached the Accident Compensation Corporation directly. It provided us with copies of judgments from its own digital files, on the condition that they not be used for commercial purposes except as specifically agreed, and that access to the data always be publicly available free of charge. Gaining access to judgments in this way may not be possible in other circumstances or for the general public.¹³

53. We reviewed the ways that people access judgments in New Zealand. We concluded that judgments may be accessed from a range of sources, but none of these sources are complete: frequently judgments are missing, either intentionally or as a matter of oversight. Even if there are published criteria illustrating how decisions are selected, there is no way for external observers to assess the extent to which databases actually comply with those criteria, or whether and on what grounds a judgment has been excluded or withheld. None of the databases that we were able to locate permitted people to download judgments in large volumes using automated processes.

54. In our original 2015 research, we accessed judgments individually using the New Zealand Legal Information Institute (NZLII). We accessed the NZLII webpage for each judgment and manually downloaded it from the hyperlinks provided. We found that there were important judgments missing¹⁴ – though we do not intend this as a criticism of NZLII, which offers an invaluable service on a small budget. In the present research project, we found that government databases only provide access to limited numbers of judgments. These are frequently limited to particular time periods. For example, many databases do not provide decisions

¹³ See New Zealand Law Commission Access to Court Records (Wellington, 2006) NZLC R93 at p 13 where the Commission noted there was a need to enable access to the court record (including judgments) for bona fide research projects.
¹⁴ See also T Mijatov, W Forster, T Barraclough “Problems with access to law in personal injury disputes” (2016) 27(2) New Zealand Universities Law Review 365.
earlier than 2014. In the course of this project, we identified a further specific situation where District Court decisions about tenancy law could not be accessed by the researchers using public databases.\(^\text{15}\)

**WHY WE SHOULD ENHANCE THE AVAILABILITY OF JUDGMENTS**

**JUDGMENTS STATE THE LAW**

55. Judgments form one of the two primary sources that comprise the laws of New Zealand, the other being legislation. On principle, the rule of law requires that people have the opportunity to inform themselves of what the law requires of them. It is a matter of constitutional principle in New Zealand that law should be freely available.\(^\text{16}\) The importance of comprehensive access to the law is undisputed.\(^\text{17}\)

56. By contrast, much less emphasis has been placed on public access to case law in the form of written judgments. Despite this, judgments are binding records of judicial interpretations of legislation. They resolve situations where statutes or statutory provisions may conflict, and chronicle the incremental development of “judge-made” common law. Moreover, our previous research has shown that, in aggregate, these judgments are a valuable source of information about how the legal system is operating.

57. In this report we refer to “case law”, “written decisions”, and “judgments” interchangeably (at times we also include written orders of the Court of general public effect, such as suppression or non-publication orders). While there are differences between these, they have in common the

\(^{15}\) See Appendix Four: Missing case law on Tenancy.

\(^{16}\) This is also an expectation of New Zealand’s trade partners and organisations like Transparency International. See information about work to make secondary legislation publicly available in machine readable formats in the Open Government Action Plan: <www.opengovpartnership.org/members/new-zealand/commitments/nz0015/>. This work is linked in part the Trans-Pacific Partnership Areement’s transparency publication requirements, and the work to make secondary legislation available, as discussed in this briefing to the incoming Attorney-General from 2017: <https://www.beehive.govt.nz/sites/default/files/2017-12/Parliamentary%20Counsel%20Office.pdf> accessed 3 November 2020.

\(^{17}\) Some jurists such argue that a system of law which relies on law that is not publicly available “is not properly called a legal system at all ...” For example, Lon Fuller’s principles of legality in Fuller, L. (1969). The Morality of Law: Revised Edition. NEW HAVEN; LONDON: Yale University Press at 39: “Certainly there can be no rational ground for asserting that a man can have a moral obligation to obey a legal rule that does not exist, or is kept secret from him ...”.
quality of being written records of the factual findings, reasons, and legal analysis followed by a judge or a statutorily empowered decision-maker like a tribunal in reaching a decision. They also record important information such as the identity of the parties to a dispute, and what they must do in order to comply with the law.

58. These records and the details they include are essential for the principle of stare decisis, or the principle that the Court should follow the reasoning of precedent set in previous cases unless the facts of the case before it are materially different. Written decisions are one of the ways that judges (and other decision-makers) justify whether they are obliged to follow a precedent. They also help people involved in or anticipating a dispute to assess how their current situation will be dealt with if it comes before the Courts. Judgments therefore play a crucial role in generating legal certainty and opportunities for dispute prevention.

59. Finally, judgments contribute to the effective administration of justice – a core principle of our legal system. It is a duty of every lawyer to facilitate the administration of justice.\textsuperscript{18} Facilitating the administration of justice is also one of the principal purposes of the Contempt of Court Act 2019, which clarifies some of the powers available to judges to ensure the integrity of the justice system in situations where, for example, suppression orders are breached.\textsuperscript{19}

**CONSTITUTIONAL FACTORS, OPEN JUSTICE AND FREEDOM OF EXPRESSION**

60. Written decisions are the primary output of the judicial arm of government in New Zealand. The judiciary has implied and inherent powers to control its own procedures\textsuperscript{20} and it forms one of the three core institutions that comprise our government under the separation of powers doctrine.\textsuperscript{21} Analysing written judgments is one of the few ways people have to assess how the judiciary is exercising its constitutional authority.

---

\textsuperscript{18} Lawyers and Conveyancers Act 2006, s 4.
\textsuperscript{19} Contempt of Court Act 2019, s 3.
\textsuperscript{20} Some courts and tribunals have less power to do this, or a heavier emphasis on statutory powers. Here again there may be reason for distinguishing between higher and lower courts and tribunals, although only on a case by case basis.
\textsuperscript{21} We note here that, at this constitutional level, there may be good reason for distinguishing between courts and tribunals based on their significance and seniority, but even low level tribunals wield significant legal power.
Ensuring that written judgments are publicly available is important for upholding the principle of open justice, explained as follows in a report by the New Zealand Law Commission:\textsuperscript{22}

The principle of open justice is a long-standing buttress of legitimate court systems, and is fundamental to New Zealand’s system of justice. It underpins the public’s right to attend court hearings, the media’s right to report proceedings, and access to court documents. Throughout history criminal proceedings have usually taken place in public and involved degrees of public participation.

Open processes are central to maintaining public confidence in the administration of justice and ensuring the accountability of judges. Openness also enhances the accuracy of the process and serves an educative function. It contributes to democratic process as the actions of Executive Government or of officials are subject to public scrutiny in the context of court proceedings.

The principle of open justice means that courts should conduct their business in the open. Any person is generally entitled to enter a court in New Zealand and observe what is going on, unless there are exceptional circumstances. This is a crucial check and balance against the powers the courts have to limit the rights of New Zealand citizens.

Moreover, the principle of freedom of expression is closely related to open justice. It dictates that people must be able to tell others what is happening in courts, not just that they should be able to go and see for themselves. In general, section 14 of the New Zealand Bill of Rights Act 1990 provides that everyone has the right to seek, receive, and impart information and opinions of any kind in any form. Again, while this right is subject to various limitations (that can be controlled by the courts), it is reasonable that it should extend to a general right to access the primary sources of the laws of New Zealand and to express opinions about them.

In a report by the New Zealand Law Commission, the Commission records a submission by the Chief Justice and the Chief District Court Judge that “that the open justice principle is satisfied by open court hearings and judgments being accessible as of right” in support of the proposition that “open justice (and consequently the accountability of the judicial process) is largely not engaged when it is a question of access to court records.”\textsuperscript{23} From this, we take it that the judiciary would agree that access to judgments is an essential component of giving effect to the principle of open justice.

\textsuperscript{23} New Zealand Law Commission Access to Court Records (Wellington, 2006) NZLC R93 at p 40.
In summary, these principles support the public's ability to scrutinise the reasons given by judges for the decisions they make. New Zealand's judges are not elected by popular vote. This changes public expectations about whether their conduct of proceedings and analysis of the law is sound, and whether it is in keeping with the dynamic expectations of the public. Written decisions are, therefore, the most important record of how judges are making decisions and what factors they say are relevant to their decision-making. As evidence of its importance we note that, on appeal, it is often the written decision which is examined by an appellate court to decide whether or not a lower court's conclusions are justified.

ACKNOWLEDGED LIMITATIONS ON PUBLIC ACCESS TO JUDGMENTS AT SCALE

The preceding points in favour must be weighed against a variety of limiting factors, both principled and pragmatic.

PRIVACY AND SENSITIVE CASES

Unlike legislation, case law deals with the application of the law to identifiable people and entities. As such, judgments often record factual findings based on evidence that concerns identifiable individuals, or which might be sensitive, or from witnesses who are vulnerable. When case law is published it contains true information about real people, often at the most acutely sensitive times in their lives. That is not the case when we publish legislation, except in the most uncommon of circumstances. As such, there is a constant need for caution.

MAINTAINING JUDICIAL DISCRETION AND CONTROL (INCLUDING SUPPRESSION)

A judgment records a point in time within a wider legal process which may be ongoing (for example, a criminal trial or appeal). This means that the availability of judgments, and their contents, must anticipate how they can unduly affect future proceedings. In particular, judgments can be subject to "suppression", or orders prohibiting publication of certain details, which limits the ability of others to publish information about court proceedings, including the names of witnesses and parties. That is because publication can affect the integrity of the legal process being followed or the administration of justice, for example by publishing incorrect or potentially prejudicial allegations about people involved, or by making people reluctant to participate in court processes due to public attention or privacy impacts.

24 New Zealand saw an important example of this recently when the High Court published sentencing remarks in R v Tarrant [2020] NZHC 2192 (Sentencing remarks of Mander J).
69. It is not always necessary to demonstrate that publication of a judgment influenced the fairness of judicial processes, or the capacity to identify a witness. On its own, the potential for justice to be disrupted by publication may be sufficient to secure a suppression order or to undermine faith in the fairness of a proceeding subject to appeal.

70. Judges (and tribunal members in some cases) require the flexibility and authority to balance the principles of open justice and freedom of expression against the integrity of judicial processes, fairness to the parties, and the sensitivity and vulnerability of the information and people involved in justice systems. The judiciary has statutory, implied, and inherent powers to control its own procedure, which vary depending on the particular Court or Tribunal involved.

71. The democratic principle of judicial independence is also essential. The judiciary should not be improperly influenced in ways that might affect the administration of justice. Furthermore, the judiciary has strictly limited capacity to respond to public criticism. This is an essential consideration when it comes to our recommendations that judges should be required to work within particular systems or to have their judgments treated in particular ways.

72. Finally we note that there are a range of statutory grounds for suppression of judgments and that some of these apply automatically, regardless of any order made by a judge.

**Would another system better protect these principles?**

73. We considered whether the existing digital and operational processes for preparing and distributing judgments is optimal. These processes are manually intensive and subject to the risks of human and communication error. The people working within this current system face conflicting imperatives of access and restriction, as discussed above.

74. Therefore, we concluded that a revised system which more fully utilises modern digital technologies would not only enhance access, but would also help protect against privacy and suppression breaches.

75. We outline this revised system in this report, but before doing so, we outline our understanding of how case law is currently accessed in New Zealand, outside of commercial publishers.
How is case law currently accessed outside of commercial publishers?

76. Even among legal practitioners, the operational process by which written judgments are created and published is relatively unknown. One of our core recommendations is that this system of judgment publication should be treated systematically and made more transparent to the public and to practitioners.

77. Readers may find it insightful to conduct legal research that they would ordinarily conduct in commercial platforms using only these free platforms, in order to gain some idea of their utility and shortcomings.

78. We have said that there are two kinds of access to be considered (and that accessibility from a disability perspective is affected by both):

a. basic access, in the sense of access to copies of decisions that exist, both individually and in large volumes; and

b. access in digital formats in ways that facilitate their use in digital systems.

79. New Zealanders access case law online through a range of online databases. Some of these are maintained by the Ministry of Justice. Relevant databases include:

- NZLII (the New Zealand Legal Information Institute). 25
- JDO (Judicial Decisions Online, operated by the Ministry of Justice). 26
- District Courts of New Zealand. 27
- Decisions of public interest on Courts of New Zealand, including a public email list that notifies subscribers when new decisions are released. 28
- The Ministry of Justice decision finder. 29
- Decisions that might be hosted on other websites and are indexed by search engines like Google, and accessible through a regular search engine.

---

• The New Zealand Law Society holds a publicly available list of databases for accessing case law in New Zealand, although many of these require a commercial subscription.\textsuperscript{30}

• The New Zealand Law Reports website, where individual cases from the official law reports can be accessed pursuant to a licence for $30 each.\textsuperscript{31}

80. One of the best and most complete databases is maintained by the New Zealand Legal Information Institute (NZLII).\textsuperscript{32} Despite this, NZLII receives no government funding.

81. The legal, governmental, and academic communities who most frequently use case law predominantly utilise commercial subscriptions to databases maintained by large multinational publishing houses. We suspect this is one reason why access to case law outside those databases has not been extensively considered.\textsuperscript{33} We were unable to assess how much New Zealanders pay annually for access to these databases, although one of our recommendations is that this be quantified.\textsuperscript{34} As one scholar puts it, by using these databases, government is effectively paying for its own work product, albeit with secondary content and superior organisation. While this made sense in a print world, where such organisation was difficult, much of that work can now be automated and digitised.\textsuperscript{35}

82. Across public-facing open-access databases there are some recurring features. These features make them inadequate for legal research purposes generally, and inferior by comparison to commercial products. All of them are insufficient to allow for development of AOAA services.

a. Judgments held within government databases are time-banded, whether within stipulated periods, or proceeding onwards from specific points (e.g. 2014 onwards). We are informed that the rationale for this is to protect the privacy of the parties, or to comply with other pieces of legislation such as the clean slate regime.

b. Some tribunals do not have a publicly available Ministry of Justice repository at all. Instead, they refer users to NZLII. We find this

\textsuperscript{32} See <www.nzlii.org>.
\textsuperscript{33} We make no comment on this, but in connection with our research, it was also suggested to us that many senior members of academia and the legal profession are authors with commercial relationships with the large legal publishers who receive remuneration for the secondary content they produce.
\textsuperscript{34} See Recommendations.
\textsuperscript{35} A J Martineau “Reinforcing the Infrastructure of Legal Research through Court-Authored Metadata” (2020) 112(1) Law Library Journal 5 at 28.
remarkable, particularly given that NZLII receives no government support. Other Courts and Tribunals refer users to NZLII for any judgments that fall outside the time bands covered within their databases.

c. None of the relevant legal databases are complete, though their degrees of completeness vary. In effect, users have no way of knowing how complete a database is, which judgments are missing or withheld, or the extent to which a database complies with any published selection criteria. For example, a judgment may fall within the broad selection criteria, but for some other reason still has not been published. A user has no way of learning this from the database itself. In rare instances where the user may independently know that a case is absent from the database, it can still be very difficult to locate. In two previous research projects we have been unable to locate judgments that we otherwise know exist.

d. Where judgments are subject to suppression orders then the entire judgment may be withheld, rather than redacting only the identifying particulars. Further, some databases state that even when a suppression banner is included in a judgment on a database, such judgments will only have been uploaded after suppression requirements are no longer in effect. This is a blunt method to reduce the risk of breach of suppression, but inefficient, as well as confusing for database users, who are being advised to ignore the contents of a banner placed on the judgment by a judge.

e. There is no way for a user to know how good the search functionality of a database is. This means that users must temper confidence in the conclusions they draw from the use of the database. This has practical legal implications. For instance, based solely on the use of these databases, it makes it difficult for lawyers (as officers of the Court) to confidently advise the Court that all relevant decisions have been located and put before the Court.

f. Databases require users to use specific techniques in order to perform an accurate search, for example by replacing certain keyboard characters with other characters. These tend to vary in subtle ways across different databases.

g. Some government databases have superior search functionality, suggesting better data organisation on the government end. For example, some databases can be searched by the outcome of the case or by the legal materials cited. It is not clear how the public is intended to gain access to such data at scale, other than as an augmentation to search functionality. We infer such data has been created for internal statistical and quality purposes by government agencies or the Ministry of Justice.

h. All relevant databases forbid users from downloading large volumes of judgments using scraping or data-mining techniques, or other automated processes. Judgments can only be accessed one at a
time, by manually locating the individual case (often hosted within its own webpage) and following the sequence pursuant to a download link.

i. Decisions tend to be available only in PDF, apart from NZLII, which also provides RTF documents in some cases and HTML web pages of judgments in most cases (but not all). Often, these RTF and HTML documents are produced by automated means, which leads to rates of error, or the inability to process footnotes. NZLII used to manually repair these footnotes but have been unable to continue because of pressures on staff and volunteer time. All HTML web pages are produced using the Australasian Legal Information Institute’s (AustLII) automated mark-up tool. Access is provided through NZLII’s relationship with AustLII’s platforms.

j. On some MOJ sites, users are advised that PDFs of judgments may only be viewed using Adobe Acrobat software. While this software is free to download, users must sign a software licence agreement to use it. Commercial versions of this same software enable a user to edit PDFs, meaning any user of the software already endorsed by the Ministry can render a PDF mutable.

k. PDFs downloaded from these databases do not provide access to headnotes, case summaries, or other metadata that facilitates the kind of superior user experience offered by commercial databases, or by accessing the official New Zealand Law Reports prepared on behalf of the New Zealand Council of Law Reporting by LexisNexis. Some Courts and tribunals, such as the ACC District Court and the disputes tribunal, have attempted to provide judicial indications of the subject of a dispute in various ways outside the body of the judgment.

l. Stakeholders we consulted noted that as a general rule, wherever there is any concern about suppression in a given case, the judgment is withheld until the suppression status can be manually verified. In practice, we heard a frequent means of tracking the state of suppression on a case was by judicial clerks or court staff communicating with each other between courts: even NZLII and AustLII are frequently required to directly contact a registry before being able to upload judgments with confidence. Paper files are the standard. Digital record keeping is not interoperable between Courts and Tribunals at various levels. We hear that, essentially, it is difficult even for the Court to know, purely from the face of a judgment, what its current suppression status is.

83. This state of affairs creates numerous barriers, inefficiencies and risks. We acknowledge that it is an improvement on the state of the system as it was five or ten years ago, however there is every reason to improve the system further.
PDF AND “ACCESS” AS MACHINE READABILITY

Judgments from official sources (i.e. the Ministry of Justice) in New Zealand are usually accessible only in Portable Document Format (PDF).

PDF is a technical standard intended to preserve the formatting and appearance of a document across various devices and various kinds of software. This makes it desirable because it mimics the way that a page of printed text functions. PDF is essentially an image format: a picture that appears like a page of text.

Anyone who has tried to copy and paste text from a PDF, especially involving footnotes or page numbers, will be aware of PDF’s limitations. Some kinds of PDFs are prepared in ways that make it easier to access the text therein, but it is often impossible for a computer to do so by default and much of OpenLaw NZ’s pipeline is devoted to text extraction.

Moreover, PDF can create accessibility barriers for people with impairments who rely on assistive technologies which are frequently incompatible with PDF. This has the effect of potentially excluding parts of the New Zealand population from accessing primary legal materials.

PDFs can be enhanced to make them more useful for digital systems and more accessible for people with disabilities. The methods of achieving this are often similar to those suggested in this report. The features that make written decisions more machine readable for wider AOAA systems are the same features that are applied for accessibility purposes, albeit with a difference in granularity and emphasis.

There is a mistaken belief in the immutability of PDF files that explains much of their appeal within the legal system. A common rationale is that the contents or a PDF are immutable, however this is incorrect. PDF documents can be easily altered using modern commercial software, such as Adobe Acrobat. As such, one of the lingering reasons for reliance on PDF in the legal system is outmoded. By contrast, using the kind of system we suggest in this report, better version control and authoritativeness could be achieved, as well as making judgments immutable in the way that many think PDF does now.

In recent years, improvements in optical character recognition ("OCR") software have increased the capacity to extract information from PDF, albeit while requiring additional processing. OCR uses artificial intelligence techniques to recognise the shapes and symbols that comprise writing and convert them into digital characters that a computer can interact with. OCR can be very good; for instance, the OCR used by OpenLaw NZ can even detect and process handwriting. However,

errors persist when large volumes of PDF files are converted to digital
text using OCR. Once extracted, it is still necessary to design algorithms
that enable computers to detect legally significant information, such as
citations to case law or legislation.

There are other features of PDFs and their legacy in paper documents
that make them undesirable for large scale computer processing. Paper
documents tend to be organised by the use of page numbers, reflecting
their historic compilation into bound volumes. Page numbers have lesser
utility when information is stored in digital systems, where information
can be rapidly located because of automated indexing. Another example
of a paper-based anachronism causing difficulty is footnotes: without
specific programming, computers fail to recognise the significance of
footnotes from PDF images and to link them to body text. Equally,
because of the constraints of paper-based publishing, citations or other
references in footnotes can be abbreviated. While this is convenient for
human authors and readers, it requires cross referencing and further
programming that adds complexity when it comes to computational
analysis.

One barrier to adopting digital publishing noted in the literature is the
absence of an agreed style guide or citation style. New Zealand has
adopted the New Zealand Law Style Guide, which formalises citation
styles and also requires that judgments be referred to by paragraph
number first, over page numbers, even in reported cases where page
numbering endures in digital systems.

Despite these limitations, PDF remains the default format for
publishing and accessing written decisions. This limits the ability for search
algorithms to process the text in a judgment, which can result in “false
negatives” when using search algorithms to locate a judgment. For
example, a case might include the word “tenancy”, but a relevant case
may not be located because the format of the judgment makes the text
of the judgment opaque to search algorithms. Even this risk of false
negatives is enough to lead practitioners to wholly reject some
databases in favour of using others.

Some organisations, such as NZLII, have developed various ways of
dealing with this problem. NZLII uses the following methods to access
the text of PDFs in computer systems:

a. a PDF converter which produces RTF files.

b. by courts emailing word documents or other formats to NZLII for
   uploading.

c. by automated HTML converters created by AustLII.

37 Peter W Martin "Abandoning Law Reports for Official Digital Case
Any situation where a PDF must be converted to another format risks generating errors. This risk can never be entirely eliminated. Subsequently, attempts to overcome one problem introduces others, with the overall effect of undermining one of the dominant rationale for distributing judgments in PDF form (that they can be reliably reproduced across various media).

Given the practical extent of our legal system’s reliance on external services like that provided by NZLII, the fact that it receives no government funding is difficult to justify. Ministry of Justice websites frequently refer citizens to NZLII as the sole means of accessing judgments outside of commercial publishers and the Ministry’s own databases. In some cases, members of the public are referred to NZLII to access any cases which fall outside of the time periods made available by the Ministry (e.g. anything before 2014). For various tribunals, NZLII offers the sole publicly available means of accessing written judgments.

As a result of what we have described, a major portion of any “pipeline” for conducting research into case law data is devoted to merely extracting useful text from digital documents that are needlessly unsuitable for computer processing, let alone actually analysing those documents for insights. This need not be the case.

In summary, written decisions today begin their lifecycle in machine readable formats. They are then then converted to formats which are not readily machine readable. For all the resources this process demands, and for all the barriers it creates, very little of value is gained. In fact, in our research and discussions with key stakeholders, we encountered no clear rationale for current processes.

**Distribution of Case Law as PDF by Email**

To our knowledge, the dominant means of distributing case law to NZLII, media, publishers, and the parties to a dispute is by attaching a PDF to an email. There are some exceptions to this, for example the recent development of JDO. There have been useful initiatives developed to improve email distribution systems for the public, including a public email list where subscribers are sent judgments of public interest from the High Court, Court of Appeal and Supreme Court.

We also understand from discussions with stakeholders that ACC – the Crown agency also acting as a party to litigation – directly loads ACC decisions of the District Court on to NZLII. We understand the Supreme Court also loads Microsoft Word copies of its decisions directly to NZLII. We are not aware of any public description of how such agreements were reached or any operational processes sitting around those mechanisms. The reliability of those approaches often relies on individual staff members, who frequently change roles, which can lead to disruption.

More importantly, it appears that each registry and Court maintains its own distribution list for judgments. It is not publicly known who is on each list, what criteria are applied to recipients or prospective applicants,
or how any new entrant to the market for legal publishing should apply to be added to the list. This could create issues of equality of access both real and perceived in terms of fairness in commercial markets and constitutional integrity. There is an important question of equity at stake: how does any organisation apply to receive judgments directly from the Court? How might any organisation apply to receive judgments in a more desirable format?

102. The system is ad hoc and iterative. A more coherent approach is desirable from the perspective of commercial, democratic and legal fairness.

**THE NEW ZEALAND COUNCIL OF LAW REPORTING**

103. The current system is best understood in its historical and technological context. Writing on paper has for some time been the ubiquitous technology of law and remained so for an extended period. The realities of paper technologies are such that they require extensive administrative and editorial attention in order to prepare for print, at which point they cannot be amended without significant effort. The optimal approach in a paper-based information system was to contract the burden of editorial work out to commercial publishers. Many of the anachronisms and idiosyncrasies of the current system are merely holdovers from the requirements of a previous system.

104. Paper-based technology also has limited applications for end-users. The only means of consuming printed text is to read it on the page with your eyes. That can be contrasted with the myriad ways that digital text can be analysed, processed, altered, hidden, indexed, and distributed in computer systems.

105. We must also acknowledge that paper-based judgments prepared by those commercial publishers, only thirty years ago, could only be accessed in hard copy. This created a practical barrier to who could access judgments and how many they could access at one time. Even five or ten years ago, it might have been possible to say that the digital nature of legal publishing systems was still emerging.

106. In New Zealand, it was important that this paper-based editorial process produced accurate and comprehensive law reports that were sufficiently curated to avoid unworkable volumes of physical written material. Indexes and digests also had to be created manually, as opposed to being an inherent feature of digital text. This system was subject to the oversight of the New Zealand Council of Law Reporting. In the digital world, those barriers are vastly reduced, and it is worth reconsidering the function of the Council. In particular, the digital capabilities available to the authors and editors of written decisions mean that there is a greater suite of editorial tools available to them and greater nuance to the way that written decisions can be published.

107. The New Zealand Council of Law Reporting is an incorporated body that was given legislative status in 1938. Its empowering statute – which was
amended in 2006 – grants a limited monopoly on the publication of reports of case law in New Zealand from the High Court, Court of Appeal and Supreme Court. The NZCLR produces the official New Zealand Law Reports and other law reports by contract with commercial publishers (such as LexisNexis).

108. Membership of the Council of Law Reporting is dictated by statute. It is comprised of: The Attorney-General; the Solicitor-General; the President of the New Zealand Law Society; a judge of the High Court nominated by the Chief Justice; and a number of members of the New Zealand Law Society as appointed by its Council. The composition of the Council is notable: it includes members of various constitutional branches of government in partnership with the legal profession. This is partly because of the potential constitutional impacts of controlling access to and curation of the cases that are said to comprise the law.

109. The rationale given for creating (and then maintaining) a limited monopoly appears to be that the quality of case reporting should not be compromised by a race to the bottom in commercial markets, or by the unnecessary duplication of a core resource for legal practitioners. Equally, it is considered important that there be official reports that are reliable, with no misreporting of important details. 38 While these rationales are important, they can be better given effect through implementing the future system proposed in this report.

110. In practice, the task of preparing official law reports appears to be contracted to a commercial publisher. Members of the public can procure individual cases from the New Zealand Law Reports from a website. 39 At the time of writing this report, each case costs $30.00 and can only be accessed using a credit card. Users are granted a licence to use the reported case and copyright is reserved by the publisher. 40 This is despite the fact that the actual judgment itself is not subject to copyright. 41 A reported case includes a head note, which gives a summary of the case as written by a lawyer. The headnote also includes subject terms and catchwords which enable indexing and organisation of case law, and are again an anachronism from printed publication days where case law digests were prepared by publishers.

39 See <www.lawreports.nz>.
41 Copyright Act 1994, s 27 (1)(g): "No copyright exists in any of the following works, whenever those works were made: ... judgments of any court or tribunal."
The NZLR website notes that members of the Supreme Court are consulted on the way Supreme Court cases are reported. A distribution sheet used by some judges asks the judge to indicate when a judgment should be published in the law reports, meaning that members of the judiciary are involved in active curation of these resources.

We have been unable to locate any public reporting about the operational activities of the Council for Law Reporting, and there are no reports we could locate on public registers.

**Conclusion to Part One**

Our research leads us to conclude that there is only incomplete and inequitable access to written judicial decisions in New Zealand. Users of commercial subscription databases have better access to the law than other members of the public in every way. This is unacceptable given the importance of judicial decisions for New Zealand’s democratic, social, economic and legal systems.

In some cases, it is only possible to publicly access decisions of courts and tribunals through NZLII, which receives no government funding and only limited governmental and judicial cooperation.

Judicial decisions are not available in formats that are suitable for digital processing, even though they are created and disseminated in digital formats. Generally speaking, they can only be downloaded one decision at a time, and cannot be accessed at scale. They are published in formats that mean text must be extracted and analysed in ways that increase the risk of error that could easily be avoided.

The system for judicial publishing is largely a result of outmoded paper-oriented publishing systems. The operation of legal publishing is opaque, ad hoc and no longer fit for purpose.

The important principles and policy considerations that both encourage and inhibit publication of written decisions could be better given effect through a digital-first system that enables mass publication of machine-readable judgments, while also giving effect to the kind of judicial control required to preserve the administration of justice and the integrity of justice systems.

---

42 The LexisNexis NZ website states: “The NZLRs maintains a high standard of accuracy. All Supreme Court decisions are reported and are approved by the Supreme Court Bench prior to publication. Supreme Court decisions are reported with counsels’ arguments (the NZLR is the only report series to provide this).” See: <www.lexisnexis.co.nz/en-nz/products/new-zealand-law-reports.page> accessed 24 August 3.50pm NZT

43 A copy of a distribution sheet was shared with us by the Ministry of Justice and is attached as Appendix Five.
PART TWO

118. In Part Two of this report, we consider what the risks and benefits may be of adopting a system that prioritises open access to judgments as data in ways that facilitate their use in automated analytic systems.

BALANCING BENEFITS AND TRADE-OFFS

119. It is our opinion that the reasons in favour of developing a system for the mass-scale delivery of judgments as data to the public outweigh the reasons for not doing so. Apart from the principled reasons for adopting such a system, we also identify specific beneficial uses that would be made possible. In reaching our conclusions in favour of such a system, we also acknowledge the potential risks that must be mitigated.

120. Before outlining our proposed system in Part Three of this report, we briefly outline the benefits and risks of a system that publishes judgments as data.

THE BENEFITS AND APPLICATIONS OF AOAA AND JUDGMENTS AS DATA

121. This report advocates for systemic changes that will facilitate the development of AOAA platforms now and in the future. At the same time, some key decision-makers may be unaware of the variety of benefits that such platforms can offer. What follows are examples that we argue can be derived from the development of judgments as data and automated open access analytics.

122. All of these examples flow from the ability to treat case law as a source of textual data for digital processing, which has been annotated using standardised metadata, thereby enabling people to use a computer to process that data in useful ways.

123. To some extent, by offering these illustrative examples here, we are giving only a narrow impression of the massive potential for human innovation that might be possible once case law is available as digital data. The potential implications of such a state of affairs are difficult to anticipate now and this is why coherent, proactive policy development is required.

EFFECTIVE PUBLIC DATABASES OF CASE LAW

124. If case law is published in standardised machine readable ways, then public databases can provide more reliable and enhanced search functionality. This will enable better quality research and increase public uptake of those platforms, rather than forcing people to rely on curated commercial databases.
**LINKED LEGAL MATERIALS**

125. Access to digital case law creates the ability to automatically link legal materials, such as legislation and case law, in a similar manner to that available to subscribers to commercial databases. The OpenLaw NZ plug-in for Google Chrome illustrates the potential of this kind of analytic work: users can access case law citing a legislative provision directly from the official NZ Government legislation website. The OpenLaw NZ platform also facilitates linking between cases, so long as a case is held within the OpenLaw NZ database.

126. This means researchers can follow a thread of case law directly, akin to the current user experience on commercial legal publishing platforms, or following articles on different subjects through Wikipedia, rather than having to conduct individual searches each time they want to locate a case.

**CITATORS**

127. A legal citator is a resource that lets a researcher know how a case has been treated by subsequent case law. It is an indicator of whether the case is useful or not as a guide to the Court’s view of the law, and whether it will be binding on particular courts. Such services are readily available through commercial legal research platforms.

128. AustLII, in association with others in the free access to law movement, have collaborated to create the LawCite citator. It "uses heuristics to recognise and extract from the content provided by those LIIs and other sources, references to cases from" tens of thousands of cases and journal articles. Because of its use of automated text analysis, the citator cannot completely remove the risk of error. If case law were prepared with this use case in mind (for example by recording where a case has been cited and whether it was cited with approval), then more accurate citators could be built by automated means.

129. It appears that data of this kind is already kept for some Ministry of Justice databases. Using the MOJ Decision Finder, users can search some databases in more granular detail using search fields related to particular legislation or case law cited, or the outcome of the case. While this data therefore exists somewhere, it has not been integrated

---


45 See for example "Deportation Decisions" available through the MOJ Decision Finder under the Immigration & Protection Tribunal, which includes fields for outcome and court decisions cited: <https://forms.justice.govt.nz/search/IPT/Deportation/>. 
into a citator style interface to our knowledge, and isn’t accessible to members of the public, or developers who might wish to attempt this.

**Tracking suppression and identifying information subject to suppression**

130. If case law is able to be analysed as a data source, then particular elements of a case can be analysed in different ways, depending on the way that computers are instructed to process that information. For example, a date can be recognised as a date and stored as a piece of data for enabling calculations. Further, if that date is tagged as being related to a particular field, for example the date a suppression order began or will be lifted, then computers can also read and understand that information. This creates possibilities for automatically tracking when suppression orders are lifted, to create greater certainty for users of case law.

131. If case law were held as structured data, then it would be possible to use automated processes to create a central electronic register of which cases have suppression orders on them. The nuanced nature of the data would enable the register to record precisely what information is suppressed, and other parameters around that suppression, including relevant dates. Other entities, such as big tech platforms and media companies, could access that register digitally and build automated systems around them to ensure specific information is dealt with appropriately.

**Research platforms for empirical work**

132. Platforms such as OpenLaw NZ can be used to algorithmically process case law and reveal interesting features across whole datasets. Other research software such as NVIVO, used by qualitative researchers for thematic analysis, can also avoid having to convert cases from PDF to other formats first. Further, particular elements of a judgment can be extracted automatically for analysis (for example, the body text without all the prefatory material). Judgments could also be filtered by factors such as outcome.

133. One method of research, “content analysis”, is a qualitative method with a long history and significant body of supporting research that enables researchers to process judgments to reach interesting findings about how the legal system functions.47

---

46 We note this is something former Minister of Justice Andrew Little asked the Ministry of Justice to explore in recent years.

One possible use of the data extracted from case law is that it can be integrated with other data, subject to acceptable controls. One example might be the integration of case law data into New Zealand’s Integrated Data Infrastructure (IDI), although significant further work would be required to do this in a durable way.

**LEGAL EDUCATION OPPORTUNITIES**

If case law is accessible as a data source, students studying law, computer science, or both, can collaborate to create innovative legal technology projects. These can be used to upskill New Zealand’s future legal and digital workforces. Projects overseas have crowdsourced legal research projects, enabling students to collaborate across universities to process legal materials using computer programming techniques. The OpenLaw NZ platform and the Research Centre are already capable of supporting such uses.

This suggests the exciting prospect that lawyers will gain greater familiarity with computer science, and computer scientists a closer understanding of the law and how it works. The impact of this should not be understated given the growing interaction between digital technologies and the law and the role that legally trained people frequently have in governance roles where organisations use digital technologies.

**INNOVATION IN LEGAL TECHNOLOGY ENTERPRISES**

If case law is a data source, and lawyers and computer scientists can collaborate and share expertise, then legal applications can be developed that may have social or commercial benefits or enhance access to justice in ways that are difficult to foresee now.

**PREDICTIVE ANALYTICS AND MACHINE LEARNING ON CASE LAW**

We heard from several stakeholders across the justice system and the legal information field the following concern: that providing access to written judgments for computational analysis will facilitate the development of machine learning on case law by the world’s biggest data

Professor Genevieve Grant for bringing these articles to our attention and we note her own work in G Grant, D M Studdert "The injury brokers - an empirical profile of medical expert witnesses in personal injury litigation" (2013) 36 Melbourne University Law Review 831.

analytics companies. In their opinion, this may not be conducive to the public interest or the administration of justice.

139. Within this concern, the two primary applications are predictive analytics (in relation to outcomes of disputes), and analytic techniques for drawing inferences about the behaviours and tendencies of legal participants in the justice system (such as judges and lawyers). We briefly explain these before saying why we believe these potential uses are not a reason for refusing to proceed with our suggested system.

**Predictive Analytics on Case Law**

140. Predictive analytics is the field of using mathematics and data to make predictions. In law, one application is the exercise of predicting the outcomes of legal disputes based on the range of insights that can be derived from analysing judicial decisions as a source of textual data. This includes not only legal reasoning, but study of language, fact patterns, decision-makers’ preferences and biases, and so on. Rooted in this possibility is the concern that such predictions may be unreliable.

141. We acknowledge the ongoing effort to develop legal predictive analytical systems. These require large volumes of judgments to be analysed using automation. \(^{49}\) Essentially, they attempt to discover relationships between specific factors of a case and the likelihood of success using quantitative analysis of the law. Such factors may not be strictly limited to doctrinal reasoning, i.e. how the law applies to the facts of the case.

142. There is academic and professional caution about the reliability of legal conclusions drawn from predictive analytics. Such work risks assuming that judgments are fulsome, objective and true summaries of all conceivably relevant features of a factual scenario. Judgments do not necessarily contain sufficient information to make accurate predictions about the outcomes of future litigation. By contrast, judgments can equally be seen as retrospective justifications for why a decision-maker has reached a particular conclusion. As summarised by Hall & Wright in relation to content analysis methods:\(^{50}\)

> Analyzing the cause-and-effect relationship between legally relevant factors and the content of judicial opinions raises a serious circularity problem: the facts and reasons found in an opinion might or might not accurately describe the real world facts or the true nature of the judge's decision-making process. There is no reason to expect that appellate

\(^{49}\) An extensive history of this kind of research is outlined in M Medvedeva, M Vols, M Wieling “Using machine learning to predict decisions of the European Court of Human Rights” (2020) 28 Artificial Intelligence and Law 237. It is situated within a wider history of empirical analysis in the law, as authoritatively summarised in Epstein, L, and Martin, D A “An introduction to empirical legal research” (2014) Oxford University Press, Oxford, United Kingdom.

\(^{50}\) M A Hall, R F Wright “Systematic content analysis of judicial opinions” (2008) 96(1) California Law Review 63 at 95.
opinions should provide complete, objective, and result-neutral statements of all the facts in each case. Instead, there is every reason to think just the opposite. Therefore, content analysts must acknowledge that a judicial opinion is the judge’s story justifying the judgment. The cynical legal realist might say that the facts the judge chooses to relate are inherently selective and a biased subset of the actual facts of the case.

143. A further point to note is that many of the attempts to develop predictive models that make headlines are developed based on incomplete datasets: for example, they may omit important legal distinctions in favour of computable outcomes or exclude portions of the judgment text that would otherwise be relevant to a human researcher.\(^5\) Frequently, the headlines reported in news media may be out of step with the limitations of the research as noted by its authors.

**ANALYSING THE BEHAVIOUR OF PARTICIPANTS IN JUSTICE SYSTEM**

144. There is also a body of work that attempts to use judgments to infer conclusions about regular participants in the judicial process. For example, some studies have tried to infer information about the judge in a case based on analysis of written decisions authored by that judge. One study attempted to infer “judicial position estimates”, for example, based on the way judges wrote their decisions.\(^5\)

145. Another significant point of contention is the way that behaviour by lawyers might be assessed using machine learning models based on cases they have appeared in. This issue has been most highly publicised in France, where there have been moves by the profession and judges to ban such applications of the technology.\(^5\)

---

\(^5\) See for example in Medvedeva (2020) at part 5.1.1: “The data we provided to the machine learning program does not include the entire text of the court decision. Specifically, we have removed decisions and dissenting/concurring opinions from the texts of the judgements.”


**WHY THESE POTENTIAL USES SHOULD NOT PREVENT DEVELOPMENT OF THE SYSTEM**

146. We do not believe these potential applications should forestall developments of the kind we recommend in New Zealand. Further, we say that the best way to proactively ensure that such applications are conducted in the public interest is to investigate reform in the judgment publication system in the manner we suggest in our recommendations.

147. The first point to note is that, while difficult in New Zealand now, a sufficiently determined and well-resourced actor can conduct these kinds of analyses already. A person can obtain a number of judgments, analyse them using predictive analytical models, and form conclusions about how a particular judge will decide a particular case, or what features of a case should be emphasised to enhance the prospect of success. The reliability of these inferences will depend, among other things, on the quality and volume of the data samples that are analysed. Further, it is not always necessary to have access to large volumes of judgments for this purpose. Some studies have used only a small number of decisions to create an algorithmic model that can subsequently be applied to a large number of decisions. A person could develop their model based on a single case: while this is unlikely to be an accurate model, they may nevertheless rely on its inferences, to their fault. Many legal practitioners will have significant volumes of decisions in their possession already, collected in the ordinary course of practice, which could also act as a training data set.

148. The general point is that if judgments were made more widely available as potential data for computational analysis, this would simply level the playing field to others who do not already have access to those same resources. This should be seen as having positive democratic implications as well as enhancing the rule of law by contributing to an equality of arms between potential legal competitors in access to legal information.

149. The second point to note is that there is an important distinction to be made between: the isolated question of whether a particular application of predictive inferential techniques to case law is meritorious in a particular context; and the broader question of whether the state should be inhibiting such applications of predictive inferential techniques to case law generally. It is worth interrogating any suggestion that the state should be denying people the right to conduct such analyses of primary legal materials where there is any prospect that they might reveal findings of public interest. When such an analysis is conducted, the merits of the findings from that analysis can be dealt with at that point, by reference to how that specific analysis has been conducted. Another related point to note is that if we want analysis of this kind to be of any merit, or for there to be any body of domestic expertise, then we should be creating the conditions required for that expertise to flourish. That
means taking steps to foster it within responsible parameters by making data available.

150. The third and perhaps most important point is that, in fact, the state already regulates the use of legal predictive analytics in certain ways in some contexts. There are existing legislative and professional obligations placed upon legal practitioners. Lawyers are subject to professional obligations of conduct and client care to clients and the court not to use such information in ways that are misleading, negligent, or detrimental the administration of justice. There are severe professional consequences if a lawyer were to rely on predictive inferences without exercising reasonable care and skill.

151. It is also important to note that the Court has significant powers to hold people in contempt (recently re-cast in the Contempt of Court Act 2019), including to prevent the publication of information, to require information to be taken down from internet sites, or to take action in situations where the administration of justice or integrity of the judiciary is at stake. This provides the Court with a wide discretion to act in situations where use of such predictive research is being used to undermine the integrity or independence of the judiciary, or the administration of justice. Where researchers are conducting research through academic institutions, this could also provide a degree of accountability for those researchers through ethical oversight committees.

152. A final point to note is that written decisions are the primary way that the public and others can assess the quality of judicial decision-making and whether judges are deciding cases in ways that the public accepts. If members of the public wish to use computational processing to assist them in scrutinising the way judges are deciding cases, then there is a strong public interest in allowing that, not to mention serious potential damage to public trust and confidence if New Zealand were to try and prevent such analysis from occurring at all.

153. By way of conclusion, we also note that people in some jurisdictions in New Zealand are already conducting quantitative analysis about the proportion of cases dismissed or allowed by individual judges using mainly manual methods. Some Ministry of Justice databases enable users to filter cases based on "outcome": this suggests some quantitative monitoring of outcomes is also occurring within Government agencies. Researchers are also collecting information about, for example, the gender of Counsel appearing in cases before higher courts. We believe it is in the judiciary’s own interest to take proactive

54 Based on private correspondence with people conducting such work in the ACC jurisdiction.
55 See J Cooper, G Schumacher “Gender ratio of counsel appearing in higher courts” (September 2018, New Zealand Bar Association).
steps to increase the chances that such analyses are performed well by skilled people in a transparent manner in the public interest.

**WE SHOULD INVESTIGATE A BETTER SYSTEM**

154. While we acknowledge that the courts of New Zealand have a hierarchy in terms of the precedential value of their decisions, we do not believe that this excuses poor access to case law in the lower courts and tribunals. There is a temptation to treat different courts and tribunals as having different overall importance, and conclude that, on the whole, current access to case law is ‘good enough’. For example, while there might be broad agreement that case law access is poor in some tribunals, so long as access is generally good for the High Court, Court of Appeal and Supreme Courts through JDO, one might argue there is broadly no policy issue with access to case law at a national level. We think this temptation must be resisted for the following reasons.

155. First, the rule of law and individual human rights to access to justice are assessed on an individual basis. People are presumed to have access to the law and to be aware of its contents. If, in even one case, a party does not have access to the law, that confers an advantage on the other party or the state. While there are ways to cure this, for example, by provision of case law from the other party or by the Court at the point of a hearing, we think it is unacceptable as a matter of policy and principle to accept that people should have unequal access to legal materials. We think this is an uncontroversial position.

156. Second, even in terms of basic access, many people do not have free and unrestricted access to primary materials that constitute the law of New Zealand. This is true for members of the public generally, as well as disabled people with specific impairments.

**A NOTE ON ISSUES OF RESOURCING**

157. Another point to emphasise is this: undeniably, there will be resourcing issues associated with this topic. The important thing is to be clear about when a barrier is a matter of resourcing, and when it is a matter of principle. The same must be said for any suggestion that the current system is vastly superior to what came before. There should be clear agreement on the ideal parameters of the system first, and any practical or resourcing barriers acknowledged separately.

158. That is also true for any suggestion that work of the kind we recommend attracts a lower prioritisation than other essential work in the justice system. We acknowledge that this may be the case, but if work of this kind is going to be deprioritised in favour of other commitments, then this should be clearly stated.

159. During the course of our research, we were urged not to set our expectations of the justice system too high. In fact, some advised us that, in their experience, receiving any kind of access to judgment distribution
lists should not be taken for granted or jeopardised by imposing more onerous demands.

160. In this report, we adopt the view that the public are entitled to expect a system of judgment distribution that is coherent and performs well according to reasonable public expectations. If there are practical or other barriers to meeting those expectations, they should be clearly communicated. That will enable others outside the justice system to demand that senior decision-makers provide the means of overcoming those barriers.

CONCLUSION TO PART TWO

161. Having weighed the risks and benefits, we consider that there are significant overall benefits to be gained by developing systems and practices for using judgments as data. We also consider that the risks can be mitigated and managed through the design process and careful use of digital systems.
PART THREE

CREATING AND PUBLISHING MACHINE-READABLE JUDGMENTS AS STRUCTURED DATA

In this part we outline a system to produce judgments and written decisions in a way, and in formats, that can be widely published and safely used in digital systems.

Here, our references to a "system" should not be interpreted as a software system only. Rather, we include within our "system" new organisational, operational and digital processes.

SUMMARY OF PART 3

The New Zealand government (and institutions and agents constituting its legal system) should resolve to produce written judgments that are suitable for computational processing. It is significantly easier, cheaper and more accurate to produce such judgments from the beginning of their lifecycle than it is to extract data from unsuitable judgments at a later point, after they have been published.

To achieve this, judgments should be created, held and accessed as "structured data". As such, and by means of structured metadata, the text of a judgment will be machine readable: in a form that indicates to a computer the significance of individual strings of text. Effectively, the judgments themselves will be like databases. Each judgment can be read, processed and understood by computer systems.

There is international precedent for this kind of approach, particularly in jurisdictions where there is a need for managing large volumes of different kinds of legal information between different countries. Systems of this kind are most developed in international institutions where multiple languages are used, such as the European Union and the United Nations. This may be increasingly important as New Zealand transitions toward honouring its treaty obligations and implementing a justice system that more fully adopts te reo Māori.

---

56 See for example ESTRELLA (European project for Standardised Transparent Representations in order to Extend Legal Accessibility) Deliverable 3.1 General XML format(s) for legal sources), 22 January 2007. IST-2004-027655. C Lup, F Vitali, E Francesconi, M Palmirani, R Winkels, E de Maat, A Boer, P Mascellani.
57 See for example the Akoma Ntoso for the United Nations (AKN4UN) project, as part of a wider UN Semantic Interoperability Framework: <https://www.unsystem.org/content/unsif-akn4un>.
It is important to note that the additional digital structuring applied to judgments need not affect the appearance of written decisions, or require that members of the judiciary change the way they write their judgments in significant ways. Researchers recommend the use of “markup languages” that create un-seen “wrappers” for text. HTML is one example of a markup language that structures text for internet users in ways that are useful to a human user, but have significantly enhanced functionality.

Rather than describing the system any further in the abstract, we describe it here in as much specific detail as we can. There is a summary outline of the system in an appendix to this report.

Much of the detail of this system will need to be worked out through subsequent policy and technical work in close consultation with key stakeholders. This reflects the fact that much of the detail and rationale for the current system is not publicly available and that responsibility and leadership of the current system sits among various institutions.

**Overview of how suggested system would work**

This system should be predominantly forward-focused, rather than retrospective. For any new judgment, once a case has been heard, the relevant judicial figure will reserve their decision and undertake to provide a written judgment for the parties, as is currently done in most cases of interest for this research. Oral decisions do still occur in various matters, but these are later reduced to writing, and these written versions can be incorporated into the wider publishing system.

Once a decision is reserved (or delivered orally), a written decision is formulated in the mind of the decision maker and expressed: some judges do this by verbal dictation and others write judgments directly. Our proposed publishing system can account for the fact that some judges will formulate their decisions as they write.

The judgment, once written, is then edited and reviewed. In some courts and tribunals, this will be done by a person other than the author. In others, for example lower level tribunals, authors may edit their own judgments. In some cases, judgments are also internally reviewed by other judicial figures before being issued. It would also be possible for markup to be added by the author at the time of writing using appropriate software. The significant changes to the system occur after this point, which until now is more or less the same as it currently exists.

What we are suggesting is that, at the point the judgment is edited and proofed, semantic tags be applied to particular strings of text. This is effectively metadata (or data about data). To the user (judges and their staff), this process need not look any different than simply using a word processor like Microsoft Word. Importantly, however, behind the interface that the user is operating, the text that they are creating will be annotated in ways that let a computer system understand the significance of that text to a human user. This allows it to be incorporated into databases that can be used for computational purposes.
Here is a more specific example of how this suggested system would work. Where a judgment cites a case, the citation is tagged so that a computer understands it is a legal citation. That citation and the metadata around it might be used later to generate an automatic hyperlink to the reference case itself, giving the kinds of functionality only currently available to users of a commercial product. Where a date is written, it can be tagged as a date, so that the computer understands it is a date, or even tagged as a specific kind of date (i.e., the date that suppression expires). That date can then be used to conduct automated calculations around timeframes, for example when ascertaining whether a given number of days have elapsed such that suppression or non-publication orders may no longer be effective. Where the name of a party is written, it is tagged as being the name of a party, and so on. There are already existing systems that have been developed to formalise and standardise the way this tagging is done. A formal system is necessary to ensure that systems that process legal materials are interoperable (that they can “talk to each other”).

A similar process can be used for tagging information that should not be published widely, such as the name or identifying particulars of an accused or where a person’s name and identifying particular are subject to automatic statutory suppression. We suggest counsel for the parties can actively engage with the Court on this process, consistent with a lawyer's duties to the Court, to clients and to the administration of justice. Ultimately, we think this will enable an electronic register of suppression orders to be constructed and accessed by digital platforms and news publishers.

The end product of this process will be that a decision is held in a computer system as structured data but can be viewed by appropriate users in any format they like, including as PDF or a Microsoft Word document. Judgments could be exported to any form desired so long as the relevant software system anticipates it: that includes PDF for printing, HTML for web publishing, or Microsoft Word for editing and drafting legal documents. The key thing is that the system will have moved away from plain text publishing, with no indication to computer systems about what that text means, and away from PDFs, where that text is often reduced to an image format requiring optical character recognition.

Further essential points to emphasise are these:

- “standards” already exist to formalise the way that data is structured and tagged. These standards have been developed and adopted by authoritative international institutions, such as the United Nations, institutions in the European Union, and others. There is a substantial body of legal academic research sitting behind the development of these standards. If necessary, the standards can also be modified for domestic legal settings before being adopted.

- Structured data could be held in either XML or JSON formats, both of which are pervasive throughout modern digital technologies,
including the internet. There is very little prospect of these technologies becoming obsolete. In fact, they are being incorporated in a foundational way in future semantic web technologies. The ability to export judgments to other formats also presents some safeguard against this risk.

- To an end user, the text of a judgment may look no different whatsoever. They can be used in court just as they are now, with even greater capacity to incorporate individual passages of text directly into other legal documents.

- To the author of a judgment, their work process may look no different at all. In fact, our research suggests that the existing “JDI” platform used by District Court judges and above is built on open source software that can be modified, and in fact already facilitates the kinds of modifications that would make this system possible.58

178. A system developed in this way would better protect the interests of justice, the administration of justice, and judicial control over legal information. With this in place, existing legal systems for exerting influence over the publication of written decisions would be supplemented by digital systems that redact certain information and restrict access to unredacted versions of judicial decisions.

**AN EXAMPLE OF AN EXISTING SYSTEM: AKOMA NTOSO**

179. We think it is important to realise that systems of this kind are well-conceptualised and have moved beyond the theoretical stage. One standard that has been developed for marking up legal documents is called “Akoma Ntoso”, abbreviated to AKN. AKN was developed by a team working for the United Nations Department of Economic and Social Affairs59 and is one of two key components in the UN Semantic Interoperability Framework for managing the high volume of documents produced by the UN and other organisations.

180. “The term Akoma Ntoso means “linked hearts” in the Akan language of West Africa”.60 It was originally developed in order to facilitate better

---

58 See <https://www.alfresco.com/>. Multiple LinkedIn profiles for various software programmers show that the JDI platform was developed on Alfresco between 2010 and 2013, although public information about it is difficult to obtain and these developers declined to speak with the researchers without permission from the Ministry of Justice.


60 See <https://en.wikipedia.org/wiki/Akoma_Ntoso>.
organisation of and access to legal information across African legislative bodies, but has much wider applicability geographically speaking, being explored and adopted to varying degrees in pan-European settings, the United States of America, the United Kingdom and Italy. Other standards for legal data have also been developed for compatibility with AKN.

181. The New Zealand Government has adopted AKN in its government digital standards catalogue: "Formerly known as the Government Enterprise Architecture for NZ (GEA-NZ) standards reference, the government digital standards is a catalogue of digital standards and related guidance that can be used by NZ government agencies." 61

182. AKN has been adopted as an open source open access standard by OASIS (the Organization for the Advancement of Structured Information Standards). OASIS describes itself as follows: 62

OASIS is the Organization for the Advancement of Structured Information Standards, a not-for-profit, international consortium that drives the development, convergence and adoption of open standards for the global information society. OASIS promotes industry consensus and produces worldwide standards for security, Cloud computing, Internet of Things, the Smart Grid, content technologies, emergency management, eGovernment, and many other areas. OASIS open standards offer the potential to lower cost, stimulate innovation, grow global markets, and protect the right of free choice of technology.

183. AKN is described as having the following purposes and functions and has more or less been designed specifically for the use case we suggest it is used for in New Zealand, albeit in a narrower judicial setting: 63

Providing access to primary legal materials, parliamentary works and judiciaries documents is not just a matter of giving physical or on-line access to them. “Open access” requires the information to be described and classified in a uniform and organized way so that content is structured into meaningful elements that can be read and understood by software applications, so that the content is made “machine readable” and more sophisticated applications than on-screen display are made possible.

184. AKN functions as a bibliographic and records retrieval organisation system. This means AKN could be used to indicate official copies of judgments and to track their authenticity and provenance using metadata and other kinds of functions inherent to the standard.

185. New Zealand’s Parliamentary Counsel Office (“PCO”) already uses XML in the process of legal drafting. It also uses “schematron” to validate the way statutes are drafted using XML markup, although we understand

---

62 See <https://www.oasis-open.org/org/faq>.
PCO have not progressed work into the use of AKN because of a perceived lack of value given the capabilities already achieved through use of non-AKN XML.

186. AKN is oriented towards the peculiarities of legal documents. This means it creates an explicit distinction between the roles of author and editor in the marking up of documents. This is vital when it comes to considering the importance of judicial independence in relation to judicial decisions and processes for verifying official versions of documents.

187. From our research, we note that AKN includes "elements" or tags that indicate to a computer system the following information: the jurisdiction of the case law; the docket number of the trial; the neutral citation; the name of the parties; the name of the lawyer; and the name of the judge. It also includes elements for noting different opinions and different arguments, with a view to developing semantic reasoning tools. These would be optional.

188. AKN can be applied at different levels of compliance given its richness. It would be possible to apply AKN selectively to judgments on a provisional basis.

189. OpenLaw NZ is investigating ways of automatedly applying Akoma Ntoso to judgments using its platform, however this is a significant exercise.

**SUGGESTED SYSTEM CAN OVERCOME CONTEMPORARY ISSUES WITH LEGAL PUBLISHING**

190. We note that a system structured in this way would avoid existing operational problems in the justice system that have little to do with otherwise facilitating AOAA, such as tracking suppression.

**SUPPRESSION AND NON-PUBLICATION ORDERS**

191. A system of this kind would enable suppression orders to be embedded into the metadata of a document. The author or editor of a judgment would be able to specifically tag the identifying particulars that are suppressed in the body of a judgment. They could even indicate the temporal limitations of that suppression in a way that would be amenable to processing as data in computer systems. We believe that this would effectively allow judgments to be automatically redacted or unredacted depending on when they are being viewed and by whom.

192. A key benefit of the system is that it would require those closest to the case – the judge, the parties and their representatives – to take ownership and direction over how the written judgment in the case is distributed. At present, much of the work in assessing what aspects of a case can be published is pushed onto publishers or onto registries who may have little to do with the original case. They are operating in an information vacuum that increases the chances of erroneous
publication. That leads to a hesitant approach that favours simply refusing to publish a judgment if there is any doubt.

193. Written decisions are the primary work product of the judiciary and of tribunal members: it makes more sense, from the perspective of effectiveness and efficiency, to simply augment the process of producing that judgment to make it more easily relied upon by others in the system.

**TAILORED ACCESS CONTROL**

194. Judgments held as structured data will be subject to greater levels of digital control over how those decisions are accessed, when, and by whom. Currently, much of the system is structured around the fact that, once a judgment is converted to PDF and distributed, it is effectively beyond control, particularly on the internet. If judgments were to be held as structured data, they could be accessed in a redacted form, or access could be restricted in ways that enable greater trust to publishers or privileged actors than to the general public, much as the current system exists now.

195. Any system of access and control would need to account for the important principle adopted by the free access to law movement (FALM) that people accessing legal materials should not be subject to unjustified surveillance.64

196. Judgments might be separated into five categories, all of which can be treated differently by digital systems in terms of how they are published:

   a. Judgments that are fully public.

   b. Judgments that are fully public, but must be redacted or removed at a particular point subsequent to publication (for example in response to the Criminal Records (Clean Slate) Act 2004, or an appeal against conviction).

   c. Judgments that are temporarily subject to non-publication or suppression orders (for example, during a trial, on an interim basis).

   d. Judgments that are subject to permanent non-publication or suppression orders (for example where they relate to a young person).

   e. Judgments that will never be published in any circumstances (rare).

197. By creating a comprehensive system, access can be controlled to particular categories in nuanced ways, or unrestricted access can be given to copies of judgments that have been confidently redacted.

Our system will also enable redaction to occur in a way that is semi-automated. For example, a judgment’s publication status could change once a particular period of time has lapsed, although a system of this kind might require close human supervision.

Similarly, a judgment may be placed into one dataset (for example the official court record or Stats NZ’s Integrated Data Infrastructure) in one form (for example (e)) and published more widely in another de-identified form (for example (d)), while retaining the original unredacted judgment as the authoritative record. This has potential benefits to be explored for the way that judgments are treated as public records under the Public Records Act 2005.

**Authoritative Versioning**

Also important is that the official versioning of a document and any amendments to it can be recorded in the metadata of the document. This would enable people to have confidence that they are accessing an authoritative version of a judgment.

Holding judgments as structured data would generate superior protections against unauthorised alteration and better evidence of when and how a judgment has been altered, if that were ever to become an issue.

**Accessibility for Users of Assistive Technologies**

When judgments are published as PDF, they are not always accessible by default for people with visual impairments using screen reading technologies. This has the effect of excluding people with disabilities from access to legal materials. This is unacceptable from a human rights and public policy perspective.

Some have pointed out that PDF formats are capable of incorporating metadata into them. They can also be prepared in ways that facilitate machine processing and use by assistive technologies. We have decided not to recommend adoption of such systems. That is because even where such a system is adopted, the text of the judgment included in the PDF will still require annotation and editing to make it useful in that way. Once that effort has been undertaken, there is no good reason for restricting publication to that one format. Instead, if that work is going to be done, a range of formats should be possible. Our pragmatic suggestion on this point should not detract from the basic expectation that disabled people are capable of gaining access to primary legal materials in accessible formats.

We note for completeness that providing information in an accessible does not mean that a particular user who wishes to access the judgement as a PDF is prevented from doing so. Structured data can be outputted as a PDF or printed onto paper if required.
“DIGITAL FIRST” CASE LAW IS A SUPERIOR SYSTEM

205. At all points within a system for creating, relying upon, and distributing written decisions, users are faced with issues about how to access that content and use it in responsible ways. At present, the system functions by having a degree of editorial intervention by judges and by the Ministry of Justice at the point a decision is written, and later when it is shared with the public, including through editorial decisions by commercial publishers, by the media, by the Law Society and by organisations such as NZLII.

206. When a PDF is distributed using a mechanism such as email, or by permitting download by a user, the sender loses all digital control over it. While there may be legal and social mechanisms for redress, those may only have retrospective effect. Similarly, the sender of the information is reliant on tech platforms to comply with any request to remove the information.

207. What we are suggesting is that judgments be “digital first”. In that way, there would be a means of ensuring that – in tightly defined circumstances set by the Court – some decisions would never be published in ways that would identify witnesses or parties or disclose sensitive information. That information could still be included in a judgment kept as the authoritative court record, but only selected parties would gain access to it in its full and unredacted form. The public could gain access to it otherwise in full. If errors are found, it is a simple process to edit and remove it, as it remains primarily within judicial control.

208. We emphasise that this system is primarily intended to proactively deal with the issues that inhibit wider publication of judgments in machine readable forms, with an overall view to increasing public access, not restricting it further.

209. Judges already write decisions to limit the inclusion of unnecessary sensitive information. We also note that some courts are consciously taking steps to write judgments differently to account for the dominant users of them: for example, the Youth Court is providing plain language summaries of judgments to account for the prevalence of neuro-disability among its defendants. The ACC District Court has also initiated a system of including an indication of the relevant issue and relevant sections of the legislation on the cover page of its judgments, or in the file name of the judgment, so that it can be easily viewed by people

browsing NZLII. We understand the Disputes Tribunal has also experimented with this approach.

**Better access to better PDFs alone is insufficient**

210. Even if comprehensive access was given to well-developed accessible PDFs, editorial issues would remain that prevent them from being widely publicised. We believe providing better access to more PDFs is not the answer to fostering development of AOAA or even providing better access to primary legal materials.

211. If an AOAA platform publicises case law that has not been editorially assessed, then it runs all of the same legal risks as a publisher of case law does now. Platforms such as AustLII are beginning to encourage others to use AustLII’s legal information for novel research, but only within the boundaries of AustLII’s own systems. This is to enable AustLII to exert control over the way its information is used and comply with its obligations as a publisher. For similar reasons, many in the free access to law movement do not permit search engines to index their repositories of case law, nor others to access large volumes of case law at scale.

212. If our suggested system were to be adopted, then the names of witnesses, parties, and key people could be automatically redacted from case law before it is published. One implication of that might be that wider access can be given to New Zealand’s redacted case law through search engines. We note that the potential for re-identification must be considered.

213. This is one reason why it would be preferable if responsibility for the editorial content of case law was adopted by the judiciary itself. Effectively, it might be possible that the only version of a case that is ever transmitted to publishers would be one where identifying particulars are redacted. In some cases, it would not even be necessary for the publisher itself to hold a local copy of the case.

214. By building a publishing system based on judgments as data, computer systems can be built around that data in reliance on it. For example, if a suppression order is imposed or lifted, that can be communicated digitally, with the scope of the order communicated digitally and implemented automatically.

---

66 See for example the inclusion of subjects in brackets in the file names on NZLII: <http://www.nzlii.org/nz/cases/NZACC/2020/>.
OTHER POLICY FACTORS TO CONSIDER

RECOMMENDATIONS IN THE UNITED KINGDOM

215. Dr Natalie Byrom has prepared a report that makes similar recommendations to those we make here. The report identifies "reforming the system for accessing primary legal data" (judgments) as being a priority need (H1).

216. At paras 4.58 and 4.59 of her report, Dr Byrom recommends that, similar to what we have attempted in this report, the current system for delivering judgments be mapped, and that machine-readable judgments be investigated for publication:

HMCTS should work with the judiciary and colleagues in the Ministry of Justice to commission an independent report which reviews the current arrangements for disseminating judgments to the public and maps the information flows from courts to publication. On the basis of this report, HMCTS and the MoJ should engage with key stakeholders to develop a publication solution that delivers free and comprehensive access to judgments in a structured machine-readable format.

217. At para 4.54, Dr Byrom explains problems with the UK system, but also recommends that machine-readable formats such as XML be used with an open XML standard (as we recommend here):

Currently the sites providing free access to judgments publish them in non-machine-readable, unstructured format. Stakeholders from lawtech and for-profit publishers reported that the failure to publish judgments in a structured, machine-readable format that defines particular elements e.g. party names, decision etc creates barriers to entry for lawtech start-ups, particularly those operating in the not-for-profit sector, as the costs associated with preparing data are prohibitive. Stakeholders argued that judgments should be made available in an open, machine-readable format (such as XML) using a consistent and open XML standard. A common set of meta-data fields should be applied and XML schema used should be capable of distinguishing the applicative/procedural part of the judgment from the mere representation of the document.

218. The UK government has published a full response and update in response to Dr Byrom’s recommendations. The recommendation is summarised on p 7 as being that:

---

67 N Byrom "Digital Justice: HMCTS data strategy and delivering access to justice - report and recommendations" (October 2019, The legal education foundation).
68 HM Courts & Tribunals Service "Making the most of HMCTS data: HMCTS’ full response and update to Dr Byrom’s recommendations" (9 October 2020) <https://assets.publishing.service.gov.uk/government/uploads/syste
HMCTS should work with the judiciary and colleagues in the MoJ to commission an independent report which reviews the current arrangements for disseminating judgments to the public and maps the information flows from courts to publication. On the basis of this report, HMCTS and the MoJ should engage with key stakeholders to develop a publication solution that delivers free and comprehensive access to judgments in a structured machine-readable format.

219. HMCTS responded by saying:

We accept the principle that more work is needed about access to judgments. This is a complex area and we will work with MoJ and the judiciary to develop proposals over the next 12 months. We expect this to be a priority area of focus for the Senior Data Governance Panel.

DEVELOPMENTS IN CANADA

220. In September 2020, a coalition of legal information providers announced the LIDI data trust. The purpose of the trust is described as being that:

LIDI operates as a steward of public, but sensitive court and tribunal rulings and other legal data. Through advanced data platforms and other technical and contractual means, LIDI takes extensive measures to ensure members and collaborators engage with the data in a manner that respects the expectations of courts and privacy officials.

221. Canada has a history of very progressive approaches to computational processing of legal materials, and a paper from 1997 suggested then that these might one day be useful in applying computational processing techniques to analyse case law.

222. We also note work being done by Jon Khan towards conscious design in justice systems. Khan’s LLM focused more widely on design and formatting standards for judicial decisions, but also reached the following conclusions in relation to organising decisions as structured data:

Instead of court staff, scholars, governments, or analytics companies laboriously restructuring the unstructured data in decisions after judges issue them, a standardized structure could ensure the data is already structured in a logical, consistent order—much like some of the better


70 See for example the way that Canada has been using an antecedent to XML to mark up decisions of its Supreme Court for more than two decades: D Poulin, G Huard, A Lavoie “The other formalisation of Law: SGML modelling and tagging” (1997) ICAIL-97 Australia.

71 Jon Khan “‘The Life of a Reserve’: How Might We Improve the Structure, Content, Accessibility, Length & Timeliness of Judicial Decisions?” (LLM Thesis, 2019, University of Toronto).
redesigned electronic medical records. Such information would be far easier to extract and digest. A standardized structure could make decisions more amenable to text-parsing, artificial intelligence, and general readability.

Algorithmic Transparency and Reliability

223. There is an important body of empirical work developing that demonstrates the way that proprietary and opaque commercial databases produced variable results for the user, both as a result of algorithmic bias and variable provider metadata practices.

224. In an article advocating for court-authored metadata, Martineau identifies a body of work by Susan Nevelow Mart\(^\text{72}\) which empirically tests the variability in results between commercially available legal research platforms.\(^\text{73}\)

Over the last several years, she has published a series of studies designed to test the effectiveness, accuracy, and consistency of the major legal research platforms. These studies have exposed great variation in the behavior of the current crop of legal research databases, which is worrisome given their veneer as objective, unbiased systems. For example, Mart uncovered significant discrepancies between Lexis and Westlaw in how each platform generates and classifies headnotes, as well as how individual headnotes link to other cases via each platform’s citator engine. Later, Mart found a great deal of variation between the results returned by Lexis Advance, Westlaw, Fastcase, Google Scholar, Ravel, and Casetext in response to relatively simple case law searches, both in terms of relevancy rankings and—more alarmingly—whether cases were included in the results at all. On a practical level, these studies showed how the algorithms and hidden metadata practices employed by research databases have concrete effects on research outcomes.

225. Notably, the OpenLaw platform is open source: the algorithms it uses are capable of being scrutinised by anyone. The same cannot be said for other databases, and it is important that the presence or absence of algorithmic transparency is included in any policy evaluation resulting from our recommendations.

Data Ethics

226. There is a growing discussion in academia, government, and public fora about the ethical obligations of people dealing in data about individuals and groups. These discussions vary depending on whether the data relates to identifiable individuals or not, but it is important to note that even de-identified data, or data that doesn’t relate to individuals at all, can


\(^{73}\) A J Martineau "Reinforcing the Infrastructure of Legal Research through Court-Authored Metadata" (2020) 112(1) Law Library Journal 5 at 12.
still raise ethical issues. In particular, de-identified data might be used to
develop algorithmic models that can still be applied to new situations in
ways that produce harmful effects. Data can also be re-identified even
where it has been anonymised, although there is some disagreement on
how far this should act as a barrier to use of such data.

Data ethics also requires researchers to consider the consequential
impacts of using data in particular ways. Again, this is a developing area,
but there should be some confidence in the fact that it is developing and
some accounting for it in any future policy work on the area of judgments
as data.

**DIGITAL INCLUSION**

We pause to note that this report is premised on accessing primary legal
materials using digital methods. In public policy, there is increasing
recognition that shifting systems to digital platforms does not mean they
become equally accessible to everyone. We do not make that
assumption here. Clearly, people without access to the preconditions for
digital inclusion will not have equal access to this system.

But they do not have adequate access to the system as it is either. One
thing that also has to be considered is the way that existing MOJ systems
require people to navigate a range of websites, adopt various tips and
tricks, understand the limitations of the databases they’re searching, and
may return false positive or false negative results. In particular, we found
that, when using the MOJ decision finder, if a user presses the “enter” key
(using Firefox on a Mac) then the page does not return any results, even
though results do exist. Instead, it requires users to use the mouse to
click the “search” button. It took us some time to work out why this was
happening. It might have entirely excluded someone already
experiencing digital exclusion.

Our recommendation is that judicial decisions be available as a kind of
base layer for a whole range of potential digital infrastructure that can sit
on top of them. Once judicial decisions are available as structured data,
that opens them up to use in innovative applications that might enhance
user access in ways that enhance digital inclusion.

**COPYRIGHT**

While there is no copyright in judicial decisions, copyright does have
some impact on our findings. There may be copyright in the code
developed to provide access to judgments as structured data or
developed by third parties to access that data. There may also be
copyright in analytic data produced about the way people are interacting
with case law, or the data produced by analysing case law using digital
processes.

We understand there is some concern among members of the judiciary
and in academic communities about the notion that case law could be
used to generate analytic data that creates substantial private value at
the expense of other matters of public interest. The first point to note is that this is currently the case already, in part because of a preferential arrangement extended to legal publishers, where they have become the primary means of accessing case law in New Zealand and have exclusive ownership over not just sophisticated analytic data, but also basic data drawn from primary case materials. The second point to note is that, if this is perceived to be a risk, then the best means of the judiciary or policymakers influencing the system in the direction of wide public interest is by embracing ownership over the system of judicial publishing. We have outlined the way that enhanced metadata and access controls can enable the judiciary to exert influence over the way information is used.

233. One point we raise for consideration is whether copyright of some kind might provide another legal mechanism for the judiciary to exercise control over judgments distributed as data, although we understand this may be treated with extreme scepticism by some.

234. The specific applications of copyright law to this future system rest heavily on how that system is eventually developed, however we say that the provisions of the Copyright Act indicate a clear parliamentary intention that access to primary legal materials must not be inhibited by any suggestion they are subject to copyright restrictions.

CONCLUSION TO PART THREE

235. It is possible to move toward this future system now using existing standards, adapting existing software and modifying the existing editorial systems for publishing judgments.
PART FOUR

236. In part four, we set out our conclusions and recommendations from our research. We then discuss several possible implementation mechanisms to undertake the policy investigation and development required and then operationalise this system.

CONCLUSIONS

237. There is an inequality of access to judicial and other written decisions as primary legal materials between commercial databases and public databases. All public databases are not complete, in the sense that they are frequently time banded, and suppressed cases, even where publishable, may not have been uploaded. There is no way to tell as a user of the system whether a case cannot be located because of an inferior search query from the user, or because of the database itself.

238. Judgments should be prepared, held and accessed as structured data. Doing so will enable them to be processed and accessed at scale using digital systems. This will enhance public access and facilitate their use as a data source for public policy purposes and for commercial applications. It will also enhance judicial control over written decisions by providing greater means of exerting digital influence as well as legal power.

239. Systems of this kind have been explored and are currently in development internationally. In particular, the UK, EU and Canada are heavily focused on improving access to judgments and the formats those judgments are available in. Multilateral institutions like the UN and the EU are particularly focused on effective information management given the sheer volume of legal information processed by those institutions.

240. To conclude, we make some recommendations. The recommendations are non-specifically directed to “New Zealand” as a whole. That is because there is a delicate constitutional balancing act in the way that judicial decisions are prepared and published which can only be resolved on a case by case basis given the task at hand. To be frank, as external researchers, it was also difficult to understand who was responsible for what aspect of the overall system of judicial publishing. Within the recommendations about New Zealand generally, we note the following.

a. A range of entities will play some role in updating access to case law in New Zealand, but we emphasise that the judicial and executive branches of the government will need to lead, and the public service to follow.

b. The judiciary is the most important actor in this space and it is the judiciary who have ultimate responsibility and lawful authority over the way their decisions are used. But the judiciary is limited in its
ability to initiate policy or operational reform in other actors. From discussions with stakeholders, there should be no expectation that the Ministry of Justice will act proactively in this field: that is a matter of constitutional propriety and deference to the judiciary.

c. The judiciary itself, understood as judges of the High Court of New Zealand, may need to be treated differently and separately from the way that other lesser courts and tribunals are treated.

d. We strongly recommend that the legal profession take a leadership role, noting lawyers' overall obligation to facilitate the administration of justice.74

RECOMMENDATIONS

241. Take a coherent and systematic approach to making the written decisions of courts and tribunals available to the public free of charge. The current system is split among multiple government sources, NZLII, and commercial publishers.

242. Consider the future of current practices which stipulate that official law reports are only available in exchange for monetary consideration through a commercial publisher. Is this consistent with our expectations of a modern justice system?

243. Map the systems by which written decisions are produced and distributed. There is public and professional interest in doing so. This can only be done with the close participation of the Ministry of Justice and the judiciary.

244. Quantify how much public money and how much private money is spent on access to judicial decisions via proprietary commercial databases, including the New Zealand Law Reports and similar products. This is essential for demonstrating the economic value of a novel system of judgment production and distribution to New Zealand. We note, however, that there are many justifications for adopting a superior system that have merit independently of their financial implications.

245. Consult with disabled people on the current accessibility to decisions of courts and tribunals. In particular, examine the effect of the way such documents are formatted and distributed. If New Zealand wishes to enable people with lived experience of disability to join the profession, it must take steps to make primary legal materials accessible. We urge caution in relation to any suggestion that accessible PDFs should be the dominant means of distribution rather than adopting our wider suggested system.

From a policy perspective, decisions by Courts and Tribunals should be dealt with separately from advisory "legal information" more generally, and from the wider Court record. Written decisions are of paramount importance when it comes to the principle of open justice and they undergo a different process than the kind of legal advisory information anticipated by the phrase "legal information" and the contested nature of documents filed with the Court and forming part of the "court record".

Take steps towards a system whereby all decisions by Courts and Tribunals are available as structured data using extensible markup language (XML) or a similar standard (such as JSON or JSON-LD). Our recommendation is that the Akoma Ntoso XML standard be investigated as it has been consciously developed for this purpose. There is a substantial body of supporting operational and academic material around the AKN standard already.

Where there is any prospect that suppression may apply in a case, decisions should be written so that identifying particulars are tagged by the authors and editors of the judgment at source, to facilitate a situation whereby the default approach is that judicial and tribunal decisions are published in a redacted form where required. The parties to a dispute can be involved in this process subject to the oversight of the judge or tribunal member. We strongly suggest that people closely involved in a case are the people best placed to consider what details might identify a vulnerable person, or otherwise require suppression.

Digital tools are not the only means available for preventing publication of suppressed information. Digital tools would be used to identify the vast majority of content, and the law and editorial responsibility would fill in the gaps that remain. The advantage for publishers is that they will have much greater access to primary materials and greater indication from the court and the parties about what information is sensitive through digital methods.

Develop a pilot program to test the new system of judicial publishing in an appropriate dispute resolution process at lower levels of the justice system, ideally a flexible but adequately resourced tribunal with sufficient volumes of decisions to enable a useful test. The Tenancy Tribunal is one candidate, particularly in light of the introduction of name suppression requirements by the government under recent statutory amendments. Once the system's performance can be demonstrated, it could be rolled out to higher level courts. In particular, we believe that the existing JDI system used by the higher courts can be modified to fulfil this function.


251. Both the government and the legal profession must consider how best to extend immediate funding to NZLII while further investigation of the wider system is undertaken. It should be a matter of some embarrassment for the government and the profession that NZLII must do so much with so little financial or other support. In other jurisdictions, a modest levy on practicing certificates is more than sufficient to fund organisations like NZLII and this might be considered by the Law Society. New Zealand should also consider whether NZLII should continue to sit within the digital and other structures of AustLII and how that has implications for what New Zealand needs from its legal information institute.

252. Consider the future of the New Zealand Council of Law Reporting in relation to any new system of legal publishing in New Zealand. Specifically, consider whether (and how) its role should expanded or reduced in order to take account of the needs of the public to access high quality, free legal information, including reported cases.

253. Computational analysis of judgments at scale is a promising method for comprehending how dispute resolution systems are operating, and how policy issues might be contributing to the formalisation of legal disputes, as well as what kinds of legal and factual issues that exist are never proceeding to dispute but should be. Judgments provide one source of much needed information about what is going on in our justice systems and are a vital data source for regulatory stewardship. New Zealand should take steps to facilitate improvements in and opportunities to conduct this kind of research.

254. There is some risk that computational analysis of judgments will result in the creation and dissemination of unreliable research findings, which are misrepresentative of conduct by judges, conduct by lawyers, conduct by litigants, or features of the dispute resolution system. There are several ways to mitigate the risk of harm from such behaviour and it should not prevent reform in the area of access to primary sources in accessible forms and formats and in sufficient volumes for the needs of users.

IMPLEMENTATION MECHANISMS

255. Our core recommendations are wide-ranging. As such, it is reasonable to anticipate challenges in initial implementation. In our opinion there are numerous policy mechanisms that could be used to gradually implement or investigate our recommendations. Notably, these sit adjacent to, or apart from, the Ministry of Justice and the court system. These mechanisms all have some overlapping influence on the focus of our

investigation (judgments as data) or the principles underlying our recommendations.

256. The Open Government Partnership could adopt, as one of its actions, an investigation into how a system of proactive openness when it comes to decisions of Courts and Tribunals as part of its upcoming Action Plan. This would be consistent with its previous focus on the availability of secondary legislation and would also have overlap with its work on documenting use of algorithmic decision-making systems in government.

257. The Government Centre for Dispute Resolution is an organisation operating out of the Ministry of Business, Innovation and Employment. It has been investigating best practice around dispute resolution in New Zealand and arrived at a set of "model standards" which can be used to measure the performance and suitability of dispute resolution systems, as well as the regulatory systems that generate those disputes. In particular, the standards require that data be collected to enable the performance of a dispute resolution system to be assessed and ensure data stewardship. If judgments were available as structured data, this would mean they can be used for this purpose, as a data source to assess how a dispute resolution system is operating and its implications for regulatory stewardship.

258. Digital accessibility standards: there is existing policy work throughout government around digital inclusion and digital accessibility standards, as well as clear digital accessibility standards that are meant to be applied by government departments (although actual compliance is suboptimal). This work is relatively diffused through government, but does represent a body of expertise with existing connections to arms of government with some responsibility for making digital information accessible. Work led by disabled people, like the Accessibility Charter also provides a lever by which insisting on digitally accessible, machine readable case law can be demanded from government.

259. Leadership by legal profession and academia: the legal profession and the Law Society carry significant influence in matters of justice policy in New Zealand. All practicing lawyers have a duty to facilitate the administration of justice and an interest in low cost access to primary legal materials through databases that are reliable and complete. Both legal academics and the legal profession are the heaviest users of commercial databases of case law: it is highly likely that they are paying a premium for what is effectively bare access to primary legal materials. Equally, lawyers have an interest in a superior system of legal publishing in terms of client protection, and taking steps to ensure that suppression and non-publication orders are effective, easy to understand and applied in digital contexts.

260. The New Zealand Council of Law Reporting is the body with statutory responsibility for law reporting in New Zealand now. It has operated with a limited monopoly since 1938. Membership of the Council includes
representatives from the profession, the judiciary, and political and apolitical members of the executive branch of government. At present, it seems that much of its task is simply contracted out to a commercial provider. It is not clear how far this system is consistent with what New Zealanders expect of a modern legal publishing system. The Council and its members are well placed to lead reform in this area or to influence reform in the public interest.

261. The Labour Government in the previous and current terms of Parliament, as well as the Minister of Social Development, have committed to a program of work directed toward Accelerating Accessibility reform. As researchers, we have participated in discussions about that work. There is no justification for adopting systems that by design exclude people with visual impairments from accessing primary legal materials. This existing work programme, and any legislative reform that results, are likely to have implications for judgment distribution. There are opportunities to progress our recommendations here as part of that work, or in anticipation of that work being completed.

262. The Parliamentary Counsel’s Office is highly experienced in the digital management and custodianship of primary legal materials. PCO has some experience in working with XML formats and has recently transitioned through a process of making online version of legislation the authoritative copies. Given the expertise PCO already holds, and the broad similarity between its present tasks and the task of making case law digital-first, PCO may be a place to go to for guidance and advice.
BIBLIOGRAPHY

Acclaim Otago Inc "Understanding the problem: An analysis of ACC appeals processes to identify barriers to access to justice for injured New Zealanders" (9 July 2015).


T J M Bench-Capon, P R S Visser "Ontologies in legal information systems - the need for explicit specifications of domain conceptualisations" (1997) ICAIL-97, Australia.

S Bruninghaus, K D Ashley "Finding factors - learning to classify case opinions under abstract fact categories" (1997) ICAIL-97, Australia.

S Bruninghaus, K D Ashley "Improving the representation of legal case texts with information extraction methods" (2001) ICAIL-2001, Missouri USA.


N Byrom "Digital Justice: HMCTS data strategy and delivering access to justice - report and recommendations" (October 2019, The legal education foundation).


P Casanovas "Legal crowdsourcing and relational law: what the semantic web can do for legal education" (2012) 5 Journal of the Australasian Law Teachers Association 159.


J G Conrad, D P Dabney "A cognitive approach to judicial opinion structure - applying domain expertise to component analysis" (2001) ICAIL-2001, Missouri, USA.


M R Dean "Independent review of the Acclaim Otago (Inc) July 2015 report into accident compensation dispute resolution processes" (May 2016)


E Francesconi "On the future of legal publishing services in the semantic web" (2018) 48(10) Future Internet


ESTRELLA (European project for Standardised Transparent Representations in order to Extend Legal Accessibility) Deliverable 3.1 General XML format(s) for legal sources), 22 January 2007. IST-2004-027655. C Lup, F Vitali, E Francesconi, M Palmirani, R Winkels, E de Maat, A Boer, P Mascellani


W Forster, T Barraclough, T Mijatov "Solving the problem: Causation, transparency and access to justice in New Zealand's personal injury system" (22 May 2017).

I Gallacher "Cite unseen: how neutral citation and America's law schools can cure our strange devotion to bibliographical orthodoxy and the constrictions of open and equal access to the law" (2007) 70 Albany Law Review 491.


M Gerami, A Hawes "Justis: at the forefront of the evolution of legal technology in the UK" (2018) 18 Legal Information Management 86


G Greenleaf, D Austin, P Chung, A Mowbray, J Matthews, M Davis "Solving the problems of finding law on the web: world law and DIAL".


High Court of New Zealand "Report from the High Court 2019 - The Year in review" (27 May 2020) (New Zealand).

P Jackson, K Al-Kofahi, A Tyrrell, A Vachher "Information extraction from case law and retrieval of prior cases" (2003) 150 Artificial Intelligence 239.

J Jenkins "What can information technology do for law?" 21(2) Harvard Journal of Law & Technology 589.


J Khan ""The Life of a Reserve": How Might We Improve the Structure, Content, Accessibility, Length & Timeliness of Judicial Decisions?" (LLM Thesis, 2019, University of Toronto).


L L Leff "Automated reasoning with Legal XML documents" (2001) ICAIL-2001, Missouri USA.


A J Martineau "Reinforcing the Infrastructure of Legal Research through Court-Authored Metadata" (2020) 112(1) Law Library Journal 5.

L T McCarty "Some arguments about legal arguments" (1997) ICAIL-97, Australia.


T Mijatov, T Barraclough, W Forster "The idea of access to justice: reflections on New Zealand’s Accident Compensation (or Personal Injury) System" (2016) 33 Windsor Y B Access Just 197


A Mowbray, P Chung, G Greenleaf "A free access, automated law citator with international scope: the LawCite project" (2016) pre-print draft.

T Neale, "Citation Analysis of Canadian Case Law" (2013) 1(1) J of Open Access to L 1.

New Zealand Law Commission "Access to Court Records" (June 2006) R93.


D Poulin, G Huard, A Lavoie "The other formalisation of Law: SGML modelling and tagging" (1997) ICAIL-97 Australia.


D Wilkins, K Pillaiapakkaminatt "The effectiveness of machine learning techniques for predicting time to case disposition" (1997) ICAIL-97, Australia.


N Terase "Is XML technology the next-generation internet communication standard? A study of leading-edge applications" PhD dissertation, 10 May 2003, Auburn University.


B Toy-Cronin, B Irvine, K Stewart, M Henaghan "The Wheels Of Justice: Understanding The Pace Of Civil High Court Cases" (November 2017, University of Otago, Dunedin, New Zealand).


R Ward "Building free access to law in Africa: some examples of successful projects" (2014) 14 Legal Information Management 290.


ABOUT THE AUTHORS

The authors were members of a research team investigating access to justice for claimants to New Zealand’s accident compensation scheme in 2015 and again in 2017. They have contributed to published articles on access to justice and two significant legal research reports on access to justice, also funded by the New Zealand Law Foundation. The authors are also members of a research team that has produced a report on how enforceable accessibility legislation for New Zealand might be designed, again with NZ Law Foundation funding.

The authors predominantly write from their experience and qualifications in law and humanities-based subjects and this is of some relevance when it comes to considering the technological aspects of our recommendations. But the authors also have experience in analysing technological subjects from this perspective, which we outline here.

Tom Barraclough is a legal researcher and policy advisor with a decade’s experience in the New Zealand legal system specialising in the areas of access to justice, environment, health, privacy, human rights, disability and technology. He is a member of the ACC Scheme Advisory Panel. With Curtis Barnes, he co-chaired the AI Forum of New Zealand’s working group on law, ethics and society.

Curtis Barnes completed an LLM with distinction on the legal and ethical questions arising from contemporary artificial intelligence paradigms. He is a founding editorial board member to Springer’s Artificial Intelligence and Ethics journal and has authored published chapters on the legal and policy impacts of computational technologies.

Mr Barnes and Mr Barraclough co-founded the Brainbox Institute, a research hub for analysing the intersection of law, policy and technology and have produced various reports on the legal impacts of synthetic media technologies. During the course of this research, Brainbox was contracted to advise the Digital Council of New Zealand on "trust and automated decision-making", another AI-related project involving close literature analysis and interviews with relevant stakeholders. Mr Barnes and Mr Barraclough are also working toward a project on "legislation as code" in New Zealand scheduled for release in 2020 with Law Foundation funding.

Warren Forster has a long history of representing people facing access to justice barriers as a lay advocate, a barrister and a researcher. He is completing New Zealand Law Foundation funded projects on the impact of neurodevelopmental impairment in the criminal justice system using data from New Zealand’s Integrated Data Infrastructure. He is also completing a proposal to reform New Zealand’s health and disability system so that it does not discriminate based on the cause of impairment. Mr Forster is a director of a company that provides alternative dispute resolution services to the Accident Compensation Corporation, called Talk Meet Resolve.
ABOUT OpenLaw NZ

OpenLaw NZ is a non-profit lawtech charity founded in 2018 by Andrew Easterbrook and William Parry in response to the lack of freely available legal materials accessible to New Zealanders. With their combined legal and technology skills, and with assistance from their volunteers, they built an open-source legal data platform to improve access to justice.

The OpenLaw NZ platform covers multiple areas including a scalable cloud pipeline to process large volumes of PDF cases into relational data, an open case law API, and a website for the public to perform their own research and improve their legal comprehension.

As part of the New Zealand Law Foundation grant that funded this project, and in conjunction with the authors of this report, OpenLaw NZ has developed a prototype Research Centre. In line with OpenLaw NZ’s vision of providing open access to high quality legal information, the Research Centre could in future be developed to provide research data directly to the public.
APPENDIX ONE:
SUMMARY OF SUGGESTED FUTURE SYSTEM

1. We propose that the authorship, editing and publication of case law be treated as a coherent system of operational, organisational and digital processes that prepare case law for use in digital systems.

2. Reform going forward should focus primarily on the creation of new judgments.
   2.1. There is a large body of case law that has already been produced in paper form, or as PDF, or as PDFs of paper documents.
   2.2. This pre-existing case law can be processed using optical character recognition and the kinds of software used by organisations like AustLII and OpenLaw NZ to create machine readable text files.
   2.3. There will be an unavoidable error rate from this transcription process, but it is scalable and automated to a degree.

3. For any new judgment, once a case has been heard, the relevant judicial figure reserves their decision and undertakes to provide a written decision for the parties.

4. The written decision is:
   4.1. Created (not necessarily in a linear fashion), by being
       4.1.1. Formulated in the mind of the decision maker;
       4.1.2. Expressed verbally by dictation, reduced to text (written or typed).
   4.2. Edited by support staff and by the author (the judge).
   4.3. Reviewed internally by the judiciary in some cases to some extent.
   4.4. The written judgment is made available:
       4.4.1. To the parties
       4.4.2. To the public (including through JDI and other platforms)
       4.4.3. To media
       4.4.4. To publishers (including commercial publishers and NZLII)
       4.4.5. To the law society
       4.4.6. To others who have been added to an distribution list.

5. As the document is created, the written decision is “marked up” by the author or the editor at first instance prior to distribution.
   5.1. Decisions can be marked up using a user-friendly interface, similar to Microsoft Word or other text editors.
5.2. The judiciary already uses a content management system called “JDI” for the preparation, sharing and distribution of judgments.

5.3. Practice varies across the Courts and Tribunals in New Zealand, although there are good public policy reasons for standardising this diversity of approach.

6. The marking up process would apply "tags" in XML (extensible markup language). We believe the Akoma Ntoso (AKN) standard should be investigated for adoption.

6.1. These XML tags can be read and understood by computer systems. The tags allow a computer system to comprehend the semantic significance of what is otherwise simply a string of text.

6.2. Without XML tags, the semantic significance of a string of text must otherwise be inferred from syntactic cues (i.e. by using heuristic techniques or RegEx).

6.3. XML tags should include anything that helps a human reader to understand the significance of different kinds of information in a judgment including, for example:

   6.3.1. Citations of legislation
   6.3.2. Citations of case law
   6.3.3. Headers
   6.3.4. Body text
   6.3.5. Paragraph numbers
   6.3.6. Footnotes.
   6.3.7. Names of parties, witnesses or other vulnerable people
   6.3.8. Dates and chronological information

7. Using XML, the author/editor must also identify and tags any details of a judgment which might be subject to suppression, for whatever reason relevant to the case.

7.1. For example, the AKN standard includes an element called “omissis”, which allows certain parts of a document to be tagged and omitted from publication by the use of software. XML elements such as this could be used to automate the redaction suppressible fields, so long as they are tagged by the author/editor of the document.

8. As generally expected already, writers of judicial decisions would adhere closely to the New Zealand Law Style Guide or other templates in order to reduce unnecessary variability in expression of core citation information, and other relevant information in the judgment, while preserving the necessary flexibility of expression within the body text.

9. The decision (and all decisions) would be made available to be accessed through a tiered system with access controls.

9.1. Where judgments are subject to suppression, access can be limited, or identifying details can be redacted from the judgment automatically.

9.2. To be consistent with rule of law principles and the declaration of the Free Access to Law Movement, users accessing this system would have to be able to do so
without being subject to surveillance. For this reason, it is likely the system would have to sit under the oversight of the judiciary, rather than the executive branch of government.

9.3. The system would be accessible by:

9.3.1. the public, to judgments in a redacted form, by querying an API or simply accessing a website where repositories of case law are held,

9.3.2. other entities with particular legal, editorial or other privileges, including researchers, the New Zealand Law Society, the news media, or trusted commercial publishers, to copies of judgments without redactions,

9.4. Functionality could be built into this system to enable traceability of individual copies of judgments that have been exported in order to notify users if a judgment’s suppression status has changed. Functionality like this would have to account for user privacy and the freedom to access legal materials without being subject to state surveillance.

9.5. Equally, embedded within any document would be metadata indicating the relevant version of that document, from an authorship perspective, allowing users to ensure they are accessing an authoritative version.

10. The exact parameters of this system, in a policy sense and a technical sense, would have to be carefully developed in consultation with stakeholders.
**APPENDIX TWO:**
**EXPLANATION OF OPENLAW NZ PROTOTYPE AND PROOF OF CONCEPT**

**OVERVIEW**

263. In this appendix, we explain how we took our previous work analysing judgments using empirical methods, and modified it to be amenable to application in the OpenLaw NZ platform.

264. In summary, we encountered a number of barriers to deploying our intended methods using computational processes. Some of those barriers arose from the way we framed our research questions. Other barriers arose because of the inherent nature of what a judgment is. Other barriers arose because of the limitations of what is possible using computational analysis of text. Other barriers were a result of inadequate access to complete databases of judgments in machine-readable formats.

265. We concluded that there is a significant opportunity for New Zealand to improve the way judgments are created and published. Not only will this have benefits for the kind of computational analysis we aimed to conduct, but there are ancillary benefits to our proposed system that we believe, of themselves, justify investigation and development.

266. Technically speaking, the prototype developed with OpenLaw NZ works as intended. We were able to analyse all 7000-odd judgments provided to us in PDF form by ACC and to draw conclusions across that entire dataset.

267. However, we found that the methods we adopted had their limitations, and further, that much of the work required by OpenLaw NZ resulted primarily from the unstructured format of judicial decisions as a data source.

**HOW WE WORKED WITH OPENLAW NZ TO DEVELOP THE Prototype**

**ENGAGEMENT WITH STAKEHOLDERS**

268. We prepared a discussion paper and circulated this to stakeholders. We had aimed to host a common discussion between the stakeholders. When this proved difficult from a logistical perspective, we conducted discussions with key stakeholders instead. We have attached the discussion paper to this report as an appendix.
REFLECTIONS ON OUR ROLE AS LEGAL RESEARCHERS WORKING WITH SOFTWARE DEVELOPERS

269. For the team of legal researchers, working with OpenLaw NZ (a software developer) was its own lesson in whether we communicated effectively, in a timely manner, and with adequate conceptual clarity and detail so as to enable OpenLaw NZ to understand the requirements of the software solution. Computers only do what they are programmed to do, and as such, refining the focus of what we wanted to search for until it was implementable by computers was a useful heuristic. It forced us to identify the ambiguities and assumptions that we otherwise took for granted. This was particularly important if research through the platform is ever going to be “crowdsourced”, as other users would inevitably end up interpreting things differently to the way that we did. Limiting the potential for vast differences between those interpretation is critical, and is primarily achieved by resolving ambiguities and assumptions.

270. The process also tested whether we really understood the nature of the task we were undertaking. As researchers with experience in arts and humanities subjects, it was easy to overlook the incredibly complex tasks that sit behind a user interface, including database structuring, writing parsers, accessing computational resources to conduct processing tasks, and even the design of the user interface itself. Frequently, we were able to mimic the language of software development, but only later would it become clear that we had no real concept of the reality behind the use of that language.

271. OpenLaw NZ were superb partners. They were passionate about the development task and the success of the project. They were truly committed to the transformative potential of their platform, the open source movement and the benefits of access to legal information for New Zealanders. Frequently, they took intelligent and cautious initiative in developing features and asking probing questions that we had not thought to outline in advance. OpenLaw NZ of course has its own significant legal expertise, and we found this invaluable.

SETTING THE SCOPE OF THE TOOL

272. We employed the following high-level steps.

---

78 We note some educational institutions are already engaged in crowdsourced tagging exercises in relation to, for example, national constitutions (the Stanford Project on Constitutional Tagging): see P Casanovas "Legal crowdsourcing and relational law: what the semantic web can do for legal education" (2012) 5 Journal of the Australasian Law Teachers Association 159.

79 We incorporated cross-checking functionality into the platform to account for this.
We reviewed the Understanding the Problem report to remind ourselves of our findings, and to assess what kinds of high-level insights we wanted to derive from the analysis of the case law.

We reviewed the documentation and tools used by the researchers to prepare the data relied upon in writing UTP, including guides prepared for the qualitative researchers and lists of questions from Survey Monkey.

We transferred all the variables from Survey Monkey into a spreadsheet using Google Sheets to enable a central, collaborative resource between the research team and OpenLaw NZ. We created a range of fields in relation to each variable, including the following.

a. "Data description" – a label for the variable, emphasising that any data intended to be captured in that variable had to be drawn from the text of the judgment itself and no other source.

b. "Researchers' priority or research value of variable" – ranked High, Medium, Low, or Not Applicable. This was intended to capture how important or significant capturing the variable would be for research purposes.

c. "Researcher's expected difficulty - to get any data or to get any good data" – we wanted to record at the outset how difficult we thought it would be to get good data in relation to that variable using computational processing.

d. "Data or funnel?" – we distinguished between what we called "funnel" parsers (which would identify passages of text that require human input to analyse) and "data" parsers such as citations (which would extract data that could be immediately entered into a database and relied upon without human input).

e. "Type of data we anticipate receiving" – we captured this so that people without legal expertise would have some idea of what kind of data the parser was meant to produce. Examples include a binary output (present or not), text, dates, citations, numbers, a frequency (i.e. number of occurrences), unique identifiers, or a selection from a pre-determined list.

f. A field with an explanation of the purpose of the data, or the desired outcome from being able to collect the data. For example, in relation to "Judge surname", we noted that we wanted to collect "number of cases heard by particular judge" and be able to conduct "cross referencing with other variables.”

g. A field for "Comments, suggestions, anticipated difficulties". This was to enable communication between OpenLaw NZ, the researchers, and anyone else. For example, in relation to "Full neutral citation", we recorded "easy if text, hard if handwriting". In relation to capturing the case number of a review decision, we recorded "These tend to follow a kind of format - usually six numerals - now perhaps seven or eight. It is probably of low importance in terms of
building a parser that can be used in non-ACC cases. Generally though, you would like to be able to link between judgments that are part of the same litigation history."

276. After a period of discussion and refinement with OpenLaw NZ, we added other columns to the sheet as follows:

a. "Revised wording of variable" – these were often framed as "judgment text includes X", where X was a key word or list of keywords. In relation to "discussion of review decision", the variable was revised to "Extract all reference to "review decision" and synonyms in judgment for human analysis".

b. "Type of data revised wording of variable will produce" – either a Boolean value, a date, an integer, a binary value, or a string of text.

c. A column to record "OpenLaw NZ’s expected difficulty" of capturing the kind of data we wanted to capture. This ranged from High, Medium, Low, Not Applicable.

277. Having refined the variables, we agreed to prioritise 5 or 6 key variables to deploy in the prototype. We elected to aim to extract all dates from a judgment and to select a short list of “funnel” variables, that would produce snippets of text for manual assessment by a human researcher.

278. For the “funnel” variables, we created a list of synonyms or keywords, which the OpenLaw NZ software would highlight for the researcher to accelerate the process of data entry.

279. Next, we worked with OpenLaw NZ, who worked with its team to design and programme the "Human Refinement Centre". This was, again, a lesson for the legal researchers in understanding the complexity and amount of work that is involved in producing a user interface that also has useful functions sitting behind it. OpenLaw NZ were highly responsive to our requests as relatively naïve users and we thank them for their patience.

280. Having tested the way the HRC worked, and suggested some improvements from a user perspective, we also tested the data export tool. This is a separate user interface which allows the user to export the data created as a result of automated processing of the judgment text, and the result of the researchers’ use of the HRC. To be specific, the tool exports data extracted automatically, such as judge name and citation information, plus any data that has been inputted by human researchers, for example, data entered to indicate whether the appeal was successful or not.

281. In future, a similar interface could be used to mark-up judgments using standardised XML in Akoma Ntoso.

**Capturing Issues with Judgments as a Data Source**

282. Within the same spreadsheet used to create variables, we aimed to collect issues faced in performing data extraction. We recorded the
workaround OpenLaw NZ used to deal with those issues, but also recorded suggestions about how judgment publishing could be modified in future to avoid those issues. Some examples of such issues are described below.

283. Some of the data on the PDF judgments we received was handwritten or inadequately scanned text. This would include either whole sentences, or situations where only one numeral in a citation was entered by hand (for example in judgment 4/2009, the 4 was handwritten).

284. This was a particular problem for older cases which had inconsistent styling, structure, and greater use of handwriting. It was necessary to develop a process to extract higher quality text content of the PDF files to enable data processing.

285. OpenLaw NZ considered various ways of dealing with this, and ultimately integrated Microsoft Azure optical character recognition into its software that was sufficiently advanced to pick up handwritten text, and improve the capture of typed text. The process provided much more accurate results, for example:

<table>
<thead>
<tr>
<th>Before (scanned PDF OCR)</th>
<th>After (Azure Cognitive Services)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The o,Str-.cT cou~T</td>
<td>IN THE DISTRICT COURT</td>
</tr>
<tr>
<td>::: Hi;L~ AT WELLINGTON</td>
<td>Decision No.</td>
</tr>
<tr>
<td>Decision N~1 /93</td>
<td>193</td>
</tr>
<tr>
<td>DCA 3/93 .:;l ...</td>
<td>HELD AT WELLINGTON</td>
</tr>
<tr>
<td>UNDER ~~~. The .. Accident</td>
<td>DCA 3/93</td>
</tr>
<tr>
<td>Rehabilitation.n.and</td>
<td>UNDER</td>
</tr>
<tr>
<td>Comp.en.sation'</td>
<td>The Accident</td>
</tr>
<tr>
<td>Insurance Act-1992</td>
<td>Rehabilitation. and</td>
</tr>
<tr>
<td>. 1:0:~~- . IN THE MATTER · of</td>
<td>Compensation</td>
</tr>
<tr>
<td>an Appeal .</td>
<td>Insurance Act: 1992</td>
</tr>
<tr>
<td>pursuant to Section 91 of the Act</td>
<td>IN THE MATTER of an Appeal</td>
</tr>
<tr>
<td></td>
<td>pursuant to Section</td>
</tr>
<tr>
<td></td>
<td>29/42/4768</td>
</tr>
<tr>
<td></td>
<td>91 of the Act</td>
</tr>
</tbody>
</table>

286. Another issue was that proprietary software is required in order to extract text from PDF files (which are essentially static images). This imposes a cost barrier to some. As a result, we recorded that judgments should be made available in text-readable format as a matter of course. OpenLaw NZ suggested HTML as one possible additional format, which led to our investigation of XML and eventual conclusions. We also noted that PDF as a format created accessibility barriers to people with visual impairments using assistive technologies.

287. Complexities arise from the judge assuming that the section being discussed in a passage of text or a footnote can be easily related by a human reader to a particular enactment. There are good reasons for this in terms of human readability and convenience, however it poses difficulties for machine readability. This was a difficult issue to deal with, because it would require imposing changes on judicial behaviour and
writing. Ultimately, we think that this will be dealt with through adoption of XML tagging, where the editing process will ensure that a reference to a bare section number also includes reference to the parent enactment.

One significant issue was to do with the variability of legal language, and the use of "legalese". We were concerned about the impact this could have on our results. One example was the variety in the ways that judges will state that an appeal has been dismissed. Based on our experience in the ACC jurisdiction (bearing in mind the number of judgments we read for UTP alone) it was easy to call to mind variable examples of judicial phrasing that might have the same substantive effect. We considered two ways of dealing with this and ultimately preferred the second. A third option was considered but judged to be too difficult within the scope of this project, providing an interesting reflection for future projects.

a. First, we could conduct a frequency analysis of the body of judgments as a whole. We could identify frequently used phrases, manually classify these, and thereby be led by the data when deciding what kinds of phrases might be relevant to particular variables. This would be a significant research task that would also be plagued by the same data extraction issues discussed above. It would also require relatively complex computational processing and data analysis. Ultimately, we elected not to pursue this option, given the scope of the present project.

b. Second, we could simply reduce the complexity of phrases we were identifying to individual words, identify those words using algorithmic processing, read them in context, and rely on our own experience to say how the data point should be captured. This approach provided a guarantee that the parser would not produce false positives or negatives, although required greater human assessment. In the event, this is what we did, and we wound up incorporating a much greater degree of human judgement into what was inferred from the presence or absence of keywords.

c. A third option, which we did not pursue, was to capture key phrases which the researcher relied upon to conclude that a variable was present or absent. Over time, this could form the basis for training machine learning models. Ultimately, this feature was raised too late in the overall trajectory of the research, and deferred for future projects. It was an example of the importance of having the legal researchers understand the way that software development works. In particular, to capture relevant passages of text that relate to variables from the judgment as a result of manual input was something that would have to result in changes at the database level of the platform. What seemed simple as an idea was technically possible and within OpenLaw NZ’s capacity, but the practical consequences of that idea were not immediately comprehended by legal researchers. If we were to conduct the project again, we would have incorporated such functionality from the outset.
Joughin v Accident Rehabilitation and Compensation...

Date of Judgment in District Court

IN THE DISTRICT COURT
HELD AT TAURANGA

IN THE MATTER of The Accident Rehabilitation and Compensation Insurance Act 1992

AND

IN THE MATTER of an Appeal pursuant to Section 95 of the Act

BETWEEN

MARGARET JOY JOUDEN

Appellant

AND

ACCIDENT REHABILITATION AND COMPENSATION CORPORATION A body corporate duly constituted under the provisions of the said Act

Respondent

HEARD on the 2nd day of August 1995

APPEARANCES

Mr N Collins for appellant
Mr A E Howman for respondent

DECISION OF JUDGE D A KENYLEY

This appeal concerns a question of interpretation of Reg 9 Accident Rehabilitation and Compensation Insurance (Decisions Costs) Regulations (No 2) 1990 SR, 1990:238.

The appellant injured her jaw and foot in a car accident on 10 August 1979. The Corporation paid some costs of treatment to her teeth. On 7 July 1989, Dr J Polske, oral and maxillofacial surgeon, wrote to the Corporation advising proposed treatment by inserting six or six left lower anterior implants to bring the appellant’s upper anterior teeth into function and to secure the line on her left lower teeth. As accompanying note from the appellant’s dentist stated:

- 2 -
EXPLAINING THE TOOL AND THE VARIABLES

289. OpenLaw NZ has created two user-facing tools as a result of this research. These tools are supported by significant software development to the OpenLaw NZ platform that are not immediately apparent to users. The two user-facing tools are:
   a. The Human Refinement Centre ("HRC")
   b. A data export tool.

290. The HRC is an interface intended to enable human researchers to work with automated processes in order to capture more complex data facets than might be possible using automated processes alone. The HRC broadly captures "funnel" variables that require human review, and the data export function captures "data" variables that can be extracted from a judgment without as much need for human review.

291. The HRC consists of the following user-facing elements:
   a. A randomiser. This takes the full body of relevant case law (ie, all ACC decisions) and enables the user to randomly sort them into sets of cases of a given number. For example, there are approximately 7000 ACC cases and the user can sort these into groups of 10 using the randomiser function. Each of these groups has a unique identifier. This enables more than one researcher to code cases and for individual researchers to be given individual sets. A pane to the left of the page lists the names of all the cases in a given set.
   b. A judgment display screen. This screen displays the entire written decision being analysed as one long scrollable pane. Keywords that have been identified as a result of automated processing are identified by coloured lines. There is also a miniature version of the pane that shows a user whereabouts they are located within the judgments as a whole (ie, in the first third, second half, etc). This is to enable a researcher to understand which "bit" of a judgment they are looking at and to quickly navigate to key bits. Within this miniature pane, keywords are highlighted, and so it functions as an overview of the case that will highlight where keywords are clustered. The judgment display screen makes navigation through a judgment much easier and faster, which was one of the key restraints encountered when manually reviewing large numbers of judgments.
   c. A list of variables that can be re-organised by clicking and dragging according to user preference: we found this was an important time-saver when analysing cases. These variables are pre-set by the researchers in partnership with OpenLaw NZ and they are the core way of creating data for answering research questions. One example of a variable is the outcome of a case: a researcher reads the case and answers the relevant question, thereby creating a
dataset about the outcome of each case analysed. We explain these variables in more detail below.

d. A set of buttons and menu options to enable data input for each variable. Depending on the variable involved, there are a set of buttons or a drop-down menu to enable the user to create data points about the case. This minimises errors in data entry, which were a significant risk in the previous 2015 research that relied on text entry into a SurveyMonkey form. The buttons are pre-set based on the requirements of the researchers and the relevant variable: for example, there are three buttons that include “Yes”, “No”, “Not applicable”, or “Unsure” to facilitate a range of situations where a variable may be present, absent, or not relevant. The “Unsure” field is captured in order to minimise the risk of incorrect data entry and to flag the data for group review later.

292. In addition to the Human Refinement Centre, OpenLaw NZ has the data export function interface. This comprises the following elements:

a. Check boxes of all relevant variables that can be selected or deselected based on user requirements.

b. A drop-down menu that enables the user to limit data extraction to particular case sets.

c. A time banding function that enables the user to limit data extraction to cases between points in time (i.e., 2010-2012 for example).

d. An additional “custom keyword” function, that can be stacked. For example, data export can include a flag for whether a particular keyword, or particular keywords exist in the case text (i.e., fields can be created for cases that include “vocational independence” and “weekly compensation”, so that analysis and processing can be restricted to only those cases where both phrases appear).

293. The data export function is used to export the data created by the user from using the HRC. It can export data that a researcher or research team has created as well as data that has been extracted using OpenLaw NZ’s parsing algorithms (or any combination).

EXPLAINING THE VARIABLES

294. For internal purposes, we distinguished between “data variables” and “funnel variables”. We are not data scientists or statisticians, so this terminology might be unorthodox.

a. Data variables refer to points of data that can be extracted from a case using automated means and more or less relied upon without further review. Examples of data variables included judge name, case citations, legislation referenced, date, etc. Many of these variables were already able to be extracted by the OpenLaw NZ platform. Some parsers (for example, ACC Case Filing Number) were specifically developed to assist with analysis of ACC cases.
b. “Funnel variables” refer to data points that require human review before being entered into the database. In the case of these variables, automated means were used to narrow down the amount of human review required by a researcher when analysing a case, but they could only go so far to reduce human involvement because of the complexity of the research question, or because of the variability in the way that relevant information in the judgment is expressed. An example is the variable ways that a judge states whether an appeal is dismissed or allowed, and on what conditions, and so “outcome” is a funnel variable, even though it might be thought of as being binary in nature.

295. Data variables in the export function include:

a. Name (of the case, particularly the two parties to the case).
b. Date of the case.
c. Citation of the case.
d. The name of the Court hearing the case.
e. A list of any case law cited in the judgment that the OpenLaw NZ platform has detected.
f. The surname of the judge hearing the case.
g. The location of the hearing as recorded in the judgment.
h. The filing number of the case if present on the judgment.
i. Whether or not the parties were represented, and if so the name of the representative.
j. A list of any legislation cited in the case that the OpenLaw NZ platform has detected.

296. It is important to bear in mind that the data variables have been automatically extracted from cases using algorithmic processing. Generally speaking, they are reliable, although it is difficult to assess their reliability without another “source of truth” beyond reviewing individual judgments themselves. The potential for false negatives or false positives resulting from this kind of automated processing is one of the strongest reasons for preparing judgments in a way that anticipates automated processing, and/or publishes judgments as structured data with semantic tagging.

297. In the export interface, the “data” variables are contrasted with what we refer to as the “funnel” variables, which are referred to under the heading User Generated Content (UGC) columns. A UGC variable can only be

---

80 While a functional version of this parser was built, OpenLaw NZ flagged that it is unstable and still in active development.
exported once at least one piece of data has been entered for that variable through the use of the Human Refinement Centre.

298. The funnel variables comprise a heading and a specific text description of the variable. These descriptions are very simple to change, although current functionality does not enable researchers to change them without OpenLaw NZ's assistance or a certain level of programming skill.

299. Funnel variables in the human researcher interface include the following, which have been defined by the research team for the purposes of testing the prototype. The answers available for each of these facets are "Yes", "No", "N/A" (not applicable), or "Unsure".

   a. Causation: "Does the judgment include discussion about any issue of medico-legal causation in the context of a claimant's access to cover and entitlements under the Act?"

   b. Absence of evidence: "Was the case decided largely on the basis of an absence of evidence (as opposed to the balance of evidence, insufficient evidence, or unpersuasive evidence), including where one party has not produced evidence relevant to the dispute before the Court?"

   c. Outcome: "As a result of the judgment, does the claimant obtain access to cover and entitlements under the Act?"

   d. Costs of litigation: "Does the judgment refer to the issue of whether the parties have incurred costs in litigating the dispute?"

   e. Jurisdiction: "Does the judgment refer to any question of whether the Court has jurisdiction (including the authority or power to hear the case, issue a judgment, or grant the relief sought by the parties)?"

   f. Review decision: "Does the judgment refer to any review application, hearing, or decision, or any finding of the reviewer (ie, pursuant to Part 5 of the Act or equivalent under previous statutes)?"

   g. Representation: "Anywhere within the judgment, does it indicate that the claimant (or other party that is not ACC) have legal representation in the case (including a non-lawyer advocate)? Does the claimant have representation?"

300. The funnel variables operate by identifying keywords in the judgment from a list defined by the researchers in relation to each funnel variable. For example, in relation to "causation", the researchers defined a list of "synonyms" or similar words for causative language, which are intended to reduce the amount of time required for a researcher to review the judgment by directing attention to passages where the synonyms appear. The human researcher is able to view this list of synonyms for each variable by clicking a small symbol next to the facet label. This feature has been added to enable the researcher to see what keywords are being searched for, and thereby assess whether a relevant keyword
has been missed by the platform for some reason, or whether it was never intended to be captured.

301. A notable result of adopting this keyword-oriented method is that it illustrates the wide variety of situations where unsophisticated algorithmic processing might have generated false positives. For example, causative language is frequently used in situations that have nothing to do with medico-legal causation or the assessment of evidence. Similarly, such language might be used in passages from evidential reports that are reproduced in the judgment, but might be subsequently rejected or ignored by the judge.

302. In addition to these funnel variables, there are also 11 date fields. These fields represent key points in the overall progression of a case’s litigation history as defined by the researchers. The intended purpose of these fields is to map the chronology of a case to inform measurements of timeliness and delay.81 The relevant date fields are, in terms of their typical chronology in a case:

a. Date of injury or health condition manifesting and requiring attention
b. Date of incapacity or date that health condition affected work
c. Date of formal claim
d. Date of most relevant ACC decision under appeal
e. Date of review application
f. Date of review hearing
g. Date of review decision
h. Date of hearing in District Court
i. Date of notice of appeal in DC
j. Date of Judgment in District Court

303. In some disputes, the date of a particular event is the thing being litigated. For example, frequently the date a claimant was unable to work, or that their health condition manifested, or that they made a formal claim will be a thing in dispute. This was why it was important to include a field for a situation where the researcher is "Unsure". In general, and in the Understanding the Problem research, we found capturing this data challenging.

304. Similarly, situations arise where multiple “decisions” by the Accident Compensation Corporation are under appeal. For this reason, we

81 Noting that timeliness cannot be assessed based solely on chronological measurements without closer contextual analysis: B Toy-Cronin, B Irvine, K Stewart, M Henaghan “The Wheels Of Justice: Understanding The Pace Of Civil High Court Cases” (November 2017, University of Otago, Dunedin, New Zealand).
deferred to researcher judgement in trying to identify the "most relevant" decision under appeal. Again, this was a challenging area and justifies further investigation.

305. The date fields function by programming the OpenLaw NZ platform to identify all text that "looks like a date", in the sense that the text is syntactically similar to the way that judges tend to write dates. We encountered some variability in the way dates were expressed: for example, initial parsers worked on a DD Month YYYY format, when in many instances, dates might be referred to as Month YYYY, or simply as a Month without any other specificity. In terms of user input, the user is able to use a drop-down menu and select any one of all of the dates identified in the case. The dates were originally arranged in that drop-down menu in the order they appeared in the case, however as a result of user feedback, they are now arranged chronologically.

306. To account for any situation where the platform misses a date such that the drop-down menu cannot be used, or where dates are expressed in an unorthodox fashion that can only be understood by a human researcher, there is also a structured text entry field that enables dates to be input in the format DD MM YYYY in three separate text boxes.

307. Next to these funnel variable facet labels, a column labelled "#" indicates to the researcher how many entries have been made for that facet in relation to that case. That lets the researcher know whether they have already made an entry for that case, or whether someone else has made an entry already. Equally, if the researcher enters data, then changes their mind or realises they have made an error, they can simply add another entry to the database. The database will automatically record that there is a discrepancy (or not) between two answers to the same variable for the same user, or two answers from two different researchers for the same variable. For each facet, consensus between users is calculated and added as a separate data item during export (for example, "jurisdiction consensus"). Consensus will evaluate true if the last added entry for a given facet is the same for each user that has entered data.

308. This enables researchers to quickly identify opportunities for “cross-checking” or “cross-coding” the way that data is being captured by different researchers, a key methodological tool for content analysis research methods. Similarly, this functionality can be used in situations where data entry is being crowd-sourced across a wide population of researchers, for example in a University environment. It would be possible to identify when a particular researcher may be entering data in ways that are out of step with other researchers. The OpenLaw NZ platform enables individual accounts to be allocated to researchers, and for data linked to those researchers to be isolated, in order to preserve data integrity.

309. The software was developed to facilitate independent review of any exported data.
OpenLaw NZ uses Git as its Source Code Management (hosted on GitHub.com), which means each code change that is committed is tracked. This is important for two reasons:

a. The code and its full development history is freely available

b. Anyone can download the code and use/inspect it for themselves

By using this foundation, OpenLaw NZ mitigates issues arising from changes to its technology over time and its impact on the dataset.

One such example of this would be:

a. Text parsers generate data that is used in a research report;

b. The extractors are upgraded (overwriting any pre-existing data); and then

c. A later researcher attempts to replicate the results of the research report.

d. It would be possible for the later user to obtain different results since the underlying extractors have been modified.

The HRC and data export address this problem by:

a. Saving the version number of the data extractors (parsers) used to extract each piece of data

b. Saving “last modified” dates for each row of data.

Thus, a person who wanted to check or validate data against its source will be able to do so, even if the OpenLaw NZ system has been significantly modified in the intervening period.

**KEY POINTS IN SUMMARY**

The following points are important to note in summary about the tool.

a. The tool is scalable. The tool enables all 7000 cases to be analysed and processed. However, the platform will process as many cases as are necessary in large volumes. With regard to automated extraction, OpenLaw NZ could extract the data required rapidly to create a system of linked legal resources.

b. During the course of the project, OpenLaw NZ has re-architected its data pipeline to perform 84 times faster than when the project began.

c. Many researchers can use the tool at once. Work can also be distributed between researchers across geographical locations. Researchers can code at the same time in parallel. It would be possible, for example, to create randomised sets of 20 or so cases, then distribute these cases across law schools in New Zealand.

d. Some data collection is automated and some requires closer human review. This is acceptable and even desirable from a research
perspective and paves the way to potential Machine Learning efficiencies in the future.

e. Once data has been created, it is stored and can be used for any research purpose required. Hence, one research project might create a dataset of which of the 7000 cases were allowed or dismissed, and another research project subsequently can use that data for further research.

f. The way that data is captured means that there is an automatic variable for “consensus”.

g. Even though “consensus” is captured, the randomiser does not automatically create overlaps between the case sets. Each case set is unique. This means any intentional cross-coding according to content analysis methods will have to occur deliberately, or as a result of regenerating case sets.

h. In practice, we have framed the funnel variables as questions, and the integrity of the data produced will rely heavily on the way these questions are framed, and how closely researchers adhere to their precise wording.

316. In hindsight, we could have sought dedicated expertise in machine learning to enhance the complexity of the parsers we were able to deploy, however our project shows that the essential data source for that kind of scholarship can be produced from PDFs using OpenLaw NZ's open source platform.

317. In hindsight, we also would have incorporated means for researchers to capture passages of text that informed their decision-making, to enable other researchers to go back and assess their analysis. Capturing such passages of text might also have created a training data set for machine learning models to further automate for relatively simple variables such as “outcome”.

LIMITATIONS OF TEXTUAL ANALYSIS OF JUDGMENTS AS A METHOD

CONTENT ANALYSIS METHODS AND JUDICIAL DECISIONS

318. We briefly note that content analysis of judicial decisions is not a new thing and has a long history in empirical legal research methods. The benefits of content analysis using judicial decisions is articulated succinctly here by Hall & Wright.82

Content analysis is a valuable research tool for more rigorous study of the empirical claims raised in conventional legal analysis. It is a more systematic and objective way to document what courts do and what

they say. The insights gained from uniform content analysis of large numbers of opinions supplement the deeper understanding of individual opinions that comes from traditional interpretive techniques. The content of judicial opinions can be important in the study of the broader social, economic, and political systems that interact with judicial precedent, but cases are also well worth scientific study in their own right.

319. In practice, we are using a mixture of human and machine methods to create dependable data about the cases in our database. We did not follow any particular methodological approach. Some of the information created about the cases is produced in a completely automated way. Most of the variables we aimed to assess from a research perspective require human input. The benefit of this method is that various data points are created about the case as a text, and later algorithmic analysis (including machine learning) could be used to assess links between the text and the data-points created by researchers.

**LIMITATIONS OF CASE LAW AS A DATA SOURCE**

320. There are several limitations to drawing inferences about the overall operation of the disputes processing system solely from judicial decisions. Again, Hall & Wright articulate those limitations succinctly.83

The major limitation of content analysis – a limit that applies equally to traditional interpretive methods – is that one cannot treat as accurate and complete the facts and reasons given in opinions. Therefore, researchers should be cautious about the meanings they attach to observations made through content analysis. This method is not well suited for predicting with assurance which cases will be won or lost. But it can tell us how cases should be developed and argued. It can also describe more accurately the landscape of decided cases, and it is the most precise way for documenting what appellate judges decide and how they explain their decisions.

321. We briefly note, in a non-exhaustive fashion, the following limitations which should be taken into account by anyone pursuing such research.

a. Judicial decisions only record features of disputes that have gone to Court. Many disputes may never reach a court or be subject to a judicial decision, but nevertheless might be important for the overall research task.

b. Judicial decisions only record features considered to be relevant by the Judge in the case and worthy of recording in the judgment. They do not record every conceivably relevant feature in a case.

c. Judgments are often written in order to be resistant to being overturned on appeal, which is not always the same as being objectively complete or accurate from a factual perspective, or as a reflection of the true causes of a dispute. For example, judges may

83 At 100.
seldom comment on matters of policy, instead preferring to focus only on the words of the statute and the evidence before them. They are not therefore intended to be a record of policy issues.

d. There is a history of respect for constitutional norms, including an attitude of deference by the judiciary toward Parliament as a democratically elected body. This may have the effect of limiting the way that issues with legislation or policy are said to have caused a dispute or perpetuated injustice within a judgment. This may mean judgments are a poor record of such things.

e. The way that a set of facts is framed into a legal dispute is not value neutral and is often strategic. For example, the same set of facts can be moulded into various causes of action or legal complaints as a result of a wide array of strategic factors. Judgments should not be analysed as if they are a naturalistic phenomenon that has occurred without human influence.

f. The Judge’s impression of the case is shaped by the issues put to the Court by the parties. These issues are not always complete and there can be reasons for excluding relevant issues or facts from judicial writing, even though they may empirically be said to have some relevance to the matter before the Court.

**LIMITATIONS OF OUR SUBJECT MATTER AND RESEARCH QUESTIONS**

322. We also note that there are limitations to our attempts to use automated and semi-automated content analysis techniques in relation to the particular subject matter of our research, access to justice.

323. Many of the things we tried to ascertain with our previous research are qualitative or even inherently legally arguable. We have tried to limit ambiguity in the way we incorporated that research exercise into this tool and a computational system was a useful limitation on the kinds of questions we could ask and answer.

324. Another point to note is that the presence or absence of data in a case is hard to assess normatively without entering into the merits of the case. For example, it cannot be said whether a complete absence of reference to case law is a flaw without forming a view on whether reference to case law was required in the context of the case.
APPENDIX THREE: NOTABLE FEATURES OF AKOMA NTOSO

325. In this appendix we note key points about the Akoma Ntoso standard identified in our research. We collate them as a foundation for future work into considering whether it should be adopted as a means of standardising the way that judgments are structured as data.

EXAMPLE OF AN EXISTING STANDARD: AKOMA NTOSO

326. A judgment is, in some ways, just a string of characters input by a human operator and stored as digital data. We perceive those characters as text, formed into words, sentences, paragraphs, pages and documents. Generally, judgments still mimic the conventions of documents intended for paper printing and distribution.

327. Without further prompting, a computer is unable to assess the significance of any of these units of meaning in a document from a human perspective.

328. "Markup languages" are used to create "wrappers" for plain text that enable a computer to comprehend the significance of a particular piece of text. For example, AKN includes a series of "tags" (or "elements") that can be used to tell a computer what a string of text means. It is a method of saying to a computer "This string of text is a citation. This string of text is the body of the judgment. This document is a judgment. This is the judge's name."

329. These wrappers can be standardised in order to facilitate development and innovation in shared information systems.

330. One standard that has been developed for marking up legal documents is called "Akoma Ntoso", abbreviated to AKN. AKN was developed by a team working for the United Nations Department of Economic and Social Affairs. Further information about AKN from a legislative perspective is available in Monica Palmirani and Fabio Vitali "Akoma-Ntoso for Legal Documents" in G Sartor et al (eds) Legislative XML for the Semantic Web Law, Governance and Technology Series 4, DOI 10.1007/978-94-007-1887-6_6 (Springer, 2011).

331. "The term Akoma Ntoso means "linked hearts" in the Akan language of West Africa". It was originally developed in order to facilitate better organisation of and access to legal information across African legislative bodies, but has much wider applicability geographically speaking, being


explored and adopted to varying degrees in pan-European settings, the United States of America, the United Kingdom and Italy. Other standards for legal data have also been developed for compatibility with AKN.

332. The New Zealand Government has adopted AKN in its government digital standards catalogue. “Formerly known as the Government Enterprise Architecture for NZ (GEA-NZ) standards reference, the government digital standards is a catalogue of digital standards and related guidance that can be used by NZ government agencies.”

333. AKN has been adopted as an open source open access standard by OASIS (the Organization for the Advancement of Structured Information Standards). OASIS describes itself as follows:

OASIS is the Organization for the Advancement of Structured Information Standards, a not-for-profit, international consortium that drives the development, convergence and adoption of open standards for the global information society. OASIS promotes industry consensus and produces worldwide standards for security, Cloud computing, Internet of Things, the Smart Grid, content technologies, emergency management, eGovernment, and many other areas. OASIS open standards offer the potential to lower cost, stimulate innovation, grow global markets, and protect the right of free choice of technology.

334. AKN is described as having the following purposes and functions and has more or less been designed specifically for the use case we suggest it is used for in New Zealand, albeit in a narrower judicial setting:

“Providing access to primary legal materials, parliamentary works and judiciaries documents is not just a matter of giving physical or on-line access to them. “Open access” requires the information to be described and classified in a uniform and organized way so that content is structured into meaningful elements that can be read and understood by software applications, so that the content is made “machine readable” and more sophisticated applications than on-screen display are made possible.

The opportunity to make visible the structures and semantic components of parliamentary, legislative and judiciary documents to software applications means to be able to use the huge capacity of ICTs to manipulate documents not as just plain undifferentiated text but in their structure and semantic components so that high value information services can be developed to assist institutions and citizens to better play their respective roles.”

---

87 See: <https://www.oasis-open.org/org/faq>.
AKN is described as having the following purpose: There are four aspects to any document that are relevant to mention, and this is also true for parliamentary, legislative and judiciary documents:

- **Content** – the actual set of words and punctuation that form the sentences of the text;
- **Presentation** – how the information looks, e.g., the colour of the text used in the document, the font used in the headings and other such formatting issues;
- **Structure** – how the information is organized, e.g., the identification of some parts of text as headings, some parts as clauses, etc.;
- **Semantics** – what the information represents or means;

Online publishing of documents has long been confined to presentation issues. Documents have been put online trying to replicate as much as possible the layout and formatting of paper. The way a document looks is very important to the "human reader" but does not really provide much useful information to the computer to actually "use" a document the way a knowledgeable human being could do.

The development of descriptive markup meta-languages such as XML allows to add [sic] information to any document that would make both the structure and the semantics of a document "usable" by a computer. Computers do not have the kind of experience and knowledge that allow professional human being [sic] to be able to deduce structure and semantics from a document unless this document has been previously "marked up" to make it "machine readable".

More specifically:

- **Structural markup** – refers to the categorizing of different parts of a text based on their role in organizing the document (e.g., sections and clauses, preambles and attachments, headings and bodies, etc.)
- **Semantic markup** – refers to the categorizing of different parts of a text based on their meaning with regard to the topic of the document (e.g., provisions, definitions, reference, names, dates, places, etc.) Sometimes structural and semantic markup overlap, e.g., in a parliamentary document a Question or a Motion is both a specific structure within the document, as well as a semantic indication of the content that will be found therein.

Akoma Ntoso provides a way to move digital documents from the presentation to the semantic era. Digital parliamentary, legislative and judiciary documents are not just displayed online, they are now "understood" by software applications and used for a myriad purposes. Both the "meaning" and "structure" of every element in a parliamentary, legislative or judiciary document are available for all applications to

---

access, thus providing the unprecedented opportunity to exploit the speed and accuracy of ICTs to manage, access and distribute such documents.

336. AKN functions as a bibliographic and records retrieval organisation system. This means AKN could be used to indicate official copies of judgments and to track their authenticity and provenance using metadata and other kinds of functions inherent to the standard.

337. AKN is a markup language that uses semantic tags. This means it can also have benefits for consumption by computers systems that have an assistive function (ie, screenreaders).

338. New Zealand’s Parliamentary Counsel Office (“PCO”) already uses XML in the process of legal drafting. It also uses “schematron” to validate the way statutes are drafted using XML markup. PCO have not progressed work into the use of AKN.

**Practical implementation of AKN**

339. AKN is oriented towards the peculiarities of legal documents. This means it creates an explicit distinction between the roles of author and editor in the marking up of documents.

340. For judges, as the authors of documents, this means that there would be little additional learning or effort involved to render their existing authorial output into the appropriate marked up form.

341. We were grateful to the Ministry of Justice and others for their discussions about the way judgments are currently prepared and held. Our research has led us to adopt the following findings:

a. Courts in New Zealand use a piece of software referred to as “JDI”. JDI is a custom build of an open source piece of content management software called “alfresco”.

b. JDI is used to publish to JDO.

c. From discussions with sales representatives for alfresco, it is clear that the platform has been built with the intention of producing content in multiple formats.

d. The platform as it is already has capacity to produce content in markup languages (including HTML).

e. As an open source platform, it could be modified to embrace AKN standards.

342. From these findings, we conclude that, without any significant changes to the way that stakeholders currently use the JDI platform, functionality could be added to enable all new judgments produced in that platform to be marked up using AKN.
From JDI, judgments could be published in a range of formats, including as structured data in AKN formatted XML, or in PDF format for printing as necessary.

Tags could be added to judgments using an ordinary user interface, such that there would be no significant difference in user experience between using Microsoft Word or any other text editor and the new judgment production interface.

The separation between author and editor would be maintained, however the role of the editor would become more specialised and enhanced. This is justifiable on the basis of the diverse advantages set out elsewhere in this report. In particular, the author and immediate editor of a judgment are those best placed to assess, with contextual knowledge of the case, which details from the judgment might identify a victim or require suppression from publication.

Much of AKN borrows elements or core principles from HTML, which is a fundamental language used in the architecture of the internet. Accordingly, there should not be any significant skill shortage beyond that already existing in the law/technology area. People frequently teach themselves to use HTML using self-directed online courses.

We note the following points from the AKN Core Vocabulary document.\textsuperscript{90}

a. AKN has been created to be customised for local jurisdictions and remain interoperable with other implementations of AKN in other jurisdictions.

b. AKN insists on careful distinction between the roles of author and editor. Semantic tags are added by an editor according to the intention of the author.

c. Typographical and other choices about the visual structure of a document (eg alignment, bolding, italics) can be reflected in the document, thereby preserving formatting from a visual perspective in a similar manner afforded by PDF (it is possible to use CSS, again, a foundational technology for the internet).

d. AKN could ultimately be used to enhance interoperability with other kinds of legal documents, including Hansard.

e. In \textit{Eight Mile Style v New Zealand National Party [2017] NZHC 2603} the Court was asked to assess copyright infringement. Its judgment at paras 10 and 11 includes hyperlinks to audio files hosted by the

Ministry of Justice. AKN would create a standardised way of linking such files to a judgment using the `<attachments>` element or similar.

f. AKN creates ways of tracking versions of documents, which is important for any situation where approval is required before publication of a judgment.

g. Information such as the result of the decision can be captured using analytical metadata in the Analysis block under the judicial field (deny, dismiss, uphold, revert, replaceOrder, remit, decide, approve). Qualifications for classifying citations can also be added. See p 51 of the AKN core vocabulary for more detail. These can include significant detail, including as to whether a citation dissents from another source.

h. AKN includes elements that indicate to a computer system the following information: the jurisdiction of the case law; the docket number of the trial; the neutral citation; the name of the parties; the name of the lawyer; the name of the judge. It also includes elements for noting different opinions and different arguments, with a view to developing semantic reasoning tools. These would be optional.

i. AKN can be applied at different levels of compliance given its richness. It would be possible to apply AKN selectively to judgments on a provisional basis.
APPENDIX FOUR: MISSING CASE LAW ON TENANCY

348. The following is an example of a situation where material case law was not available using publicly available databases. The relevant research was conducted around 18 May 2020. We include this scenario in an appendix to illustrate that there are identifiable material instances of relevant case law on commonplace issues being unavailable outside commercial databases.

349. In the course of completing this research, one of the researchers was contacted by a friend (Mark). Mark lived in a "studio room" tenancy. Mark was dissatisfied with a range of things that his landlord had done, including entry into his room without notice in advance (or at all) to replace a light bulb, onerous restrictions on the way common facilities were used, and intrusive and pervasive “house rules” attached to various surfaces throughout the property. Mark had an opportunity to move to a new living situation and wanted to pursue it, however his understanding was that he was obliged to complete the term of his fixed tenancy agreement, which was signed for a yearly term. He had approached his landlord to seek agreement to early termination of the fixed term tenancy and the landlord had not agreed, as well as demanding evidence of the need for early termination.

350. Around the same time, an article was shared in the OpenLaw NZ slack channel, which records conversations between volunteers to OpenLaw NZ. The article was published in the University of Otago student magazine, Critic Te Arohi. It detailed the relationship between a landlord and tenant in Dunedin. Notably, the tenancy tribunal found that the tenancy in question was not a fixed term tenancy, but was instead a boarding house tenancy, which had the effect of enabling the tenant to exit the tenancy with only a brief notice period (48 hours).

351. Based on that article, we attempted to undertake research using only publicly available databases to ascertain whether Mark’s tenancy might be a boarding house tenancy rather than a fixed term tenancy, using the Critic article and the Residential Tenancies Act 1986 as starting points for investigation, as a layperson might.

352. The situation turns on whether Mark lives “in a boarding house” as defined in the Act. Paragraph (b) of the definition seemed to exclude Mark’s situation:

boarding house means residential premises—
(a) containing 1 or more boarding rooms along with facilities for communal use by the tenants of the boarding house; and
(b) occupied, or intended by the landlord to be occupied, by at least 6 tenants at any one time

353. How should para (b) be applied in this case? Is the requirement for “at least 6 tenants at any one time” applied strictly, or as a matter of overall
judgement in the circumstances? The core research question was whether the requirement that there be 6 tenants in the premises was a strict requirement or one factor among a range of factors.

354. The other occupants of the other rooms in the house have separate leases with the landlord and any relationship Mark has with those tenants is incidental to his occupation of the property. The premises include at least seven tenanted bedrooms, over two floors. Each floor has around four rooms. While occupants of the ground floor have no access to the second storey, the second storey is owned by the same landlord and also consists of individual rooms tenanted to unrelated individuals. Marketing materials for tenancies from January 2021 onward exist. These marketing materials demonstrate that the rooms are mostly double rooms (with double and queen beds) and that they are marketed together even though they exist across two separate floors in the same building. The ad states in capital letters, “FIXED TERM LEASE”.

355. The search started by locating the judgment identified in the Critic article. Identifying details from that article were used to locate the relevant judgment in the publicly available tenancy tribunal orders database. Orders are only held in this database for three years. Having identified the parties, it was possible to identify the District Court case using the parties’ names.

356. The relevant case was Man Oock Holdings. That written decision summarises a case called Raffles at para 18. In relation to whether para (b) of the definition is prescriptive or not, it strongly suggests an “overall judgement” approach.

[18] In Raffles Lodge Management Ltd v Lai Judge Phillips confirmed the approach in Saunders. In that case, His Honour upheld the Tribunal’s decision that the tenancy was a boarding house tenancy. In reaching this decision the Judge placed emphasis on the nature of the property, the six rooms specifically assigned to the occupancy of particular tenants who each had their own individual tenancy agreement; the various terms of the agreement which were more consistent with the requirements of a boarding house tenancy; the ability of the landlord to have unrestricted access to the premises (other than for the specified sleeping rooms); the “house rules” both in the Agreement itself and the “courtesy notes”; and the intention of the parties. Importantly, Judge Phillips found the clause in the agreement that “the Tenant understands that this is a studio-rooms property and not a boarding house” not to be determinative, and was not supportive of the appellant’s position that it was agreed and established the intention of the parties.

357. Ultimately, it is the substance of the tenancy relationship that determines its legal status, not what the parties have agreed to call it. This more or less directly answered my question and the facts of Mark’s case, but there was no way to directly compare the facts of Mark’s case with those in Raffles without access to the decision in Raffles itself.

358. By following the citation information available only in the judgments, two cases on this issue could be easily located using publicly available
sources, but the following case law, which was cited and applied by Courts on the issue, could not be found. Note that we have used the citations as given in those judgments rather than in neutral style, to indicate what lay litigants might be working from.


d. “*Cutlers Ltd v Olivia* DC Dunedin CIV-2013-012-000410, 1 November 2013 at [21].” The reference to *Cutlers v Olivia* is an incorrect citation to *Cutlers v Thompson*. The CIV number is the same, apart from additional zeros. This only became apparent once the case was searched using Westlaw. In other databases, the absence of any results could equally have been a result of incompleteness of the database.


359. The decisions in *Saunders, Cutlers and Raffles* conclusively answered the legal question identified. They were not publicly available. Access to those cases required a commercial subscription.

360. As part of the research exercise, several publicly available free legal information resources, were also identified. Having examined the case law directly, we concluded there were material inaccuracies in them.

361. In summary, the question of whether Mark’s tenancy was a boarding house tenancy could not be answered purely based on the statute. In fact, the statute appeared to create a set of minimum prescriptive requirements, whereas the case law treated these apparently prescriptive requirements as instead being factors as part of an overall assessment.

362. In particular, laypeople might ordinarily assume that contract law obliges them to accept that any agreement that identifies itself as creating a “fixed term tenancy” has that legal status. The case law indicates that is not correct. Notably, the law of residential tenancies is intended to enable tenants and landlords to engage in legal self-help. Lawyers are prohibited from representing parties in the tribunal unless the other party consents, or the amount at stake exceeds $6000.  

91 Residential Tenancies Act 1986, s 93(2).

APPENDIX FIVE:
JUDGMENT DISTRIBUTION SHEET
# Form 2
## JUDGMENT DISTRIBUTION SHEET

<table>
<thead>
<tr>
<th>CA Number</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NZCA Number</td>
<td></td>
</tr>
<tr>
<td>Case Name</td>
<td></td>
</tr>
<tr>
<td>Lower Court Judge</td>
<td></td>
</tr>
<tr>
<td>Hearing Date</td>
<td></td>
</tr>
<tr>
<td>Court</td>
<td></td>
</tr>
<tr>
<td>Delivered By</td>
<td></td>
</tr>
<tr>
<td>Delivery Date/Time</td>
<td>at</td>
</tr>
<tr>
<td>Mode of Delivery</td>
<td></td>
</tr>
</tbody>
</table>

**Summary:** [suitable for JPI / NZLR / Media]

---

## SUPPRESSION

**PRE-TRIAL?**

**PRE-TRIAL SUPPRESSION BANNER INCLUDED?**

**ANY OTHER SUPPRESSION?**

---

## DISTRIBUTION

**GENERAL**

**(Judges, Judges' Clerks, Law Society Libraries, Solicitor-General, Secretary for Justice)**

**PARTIES ONLY**

**(Trial Judge)**

**(Rare)**

---

## JUDGMENTS OF PUBLIC INTEREST

**COURTS OF NZ INTERNET SITE:**

(“Judgments of Public Interest”)

- all civil judgments of any public interest
- all criminal appeals involving murder, manslaughter or other serious crime
- Other appeals where there has been media interest (including at first instance)

**YES** | **NO**

**TWEET:** (max 280 characters)

**YES** | **NO**

**SUMMARY:** To be published on JPI

This may be the same as the summary above but requires sign off by the Judge for publication.

**YES** | **NO**
### NEW ZEALAND LAW REPORTS

<table>
<thead>
<tr>
<th>HIGH PRIORITY</th>
<th>MEDIUM PRIORITY</th>
<th>LOW PRIORITY</th>
<th>NOT RECOMMENDED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgment</td>
<td>Judgment</td>
<td>Judgment</td>
<td>Judgment</td>
</tr>
<tr>
<td>warrants</td>
<td>which</td>
<td>which</td>
<td>which</td>
</tr>
<tr>
<td>definite</td>
<td>warrants</td>
<td>might</td>
<td>decides</td>
</tr>
<tr>
<td>consideration</td>
<td>definite</td>
<td>hold some</td>
<td>nothing of</td>
</tr>
<tr>
<td>for inclusion</td>
<td>inclusion in</td>
<td>interest for</td>
<td>interest other</td>
</tr>
<tr>
<td>in the NZLR</td>
<td>one of the</td>
<td>brief, or</td>
<td>than to the</td>
</tr>
<tr>
<td></td>
<td>available law</td>
<td>catchline</td>
<td>parties</td>
</tr>
<tr>
<td></td>
<td>reports, or</td>
<td>type,</td>
<td>themselves</td>
</tr>
<tr>
<td></td>
<td>may warrant</td>
<td>reporting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>consideration</td>
<td>but probably</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in NZLR</td>
<td>does not</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>warrant</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>reporting in</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a series of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>law reports</td>
<td></td>
</tr>
</tbody>
</table>

**JUDGMENT BELOW REPORTED?**

- **YES**
- **NO**

### SPECIALIST LAW REPORTS

- e.g. CRNZ, NZTC, PRNZ, NZFLR, ERNZ, FRNZ, ELRNZ, NZCCLR, NZConvC, NZELC, NZLR, NZRMA, NZTC, HRNZ, NZAR, NZBLC, NZCLC, TCLR, NZCPR

### PUBLICATION RESTRICTION:

- **YES**
- **NO**

### NSW INTERNATIONAL DECISIONS OF INTEREST SITE:

- **YES**
- **NO**

### BENCH BOOK:

- **YES**
- **NO**

### JDO:

- **YES**
- **YES** (After Trial)
- **NO**

### MEDIA:

- **YES**
- **NO** (Rare)

### CONFIRMATION

I confirm this judgment is ready for distribution and publication as noted above.

**Judge:**

**Date:**
APPENDIX SIX: 
DISCUSSION PAPER ON “JUDGMENTS AS DATA”
Judgments as Data in New Zealand
Discussion paper, 20 February 2020

Executive summary

We are an independent legal research team working cooperatively with OpenLaw New Zealand. Our project, funded by the New Zealand Law Foundation, aims to develop an open-source and open-access software prototype that will enable researchers to conduct large-scale analysis of judicial decisions. Using this tool, researchers will be able to propose answers to questions of value in academic, legal and policy environments.

The project requires access to large quantities of data in the form of written decisions or judgments issued by the courts and tribunals of New Zealand. It also requires that these judgments be in a machine-readable format. It can be difficult to get access to large numbers of judgments and the format of those judgments can make them unsuitable as data for processing by computers. In order for the legal community and public at large to harness the potential of judgments as a data source for quantitative and qualitative research in the future, these issues need to be resolved.

With this paper we hope to initiate a process toward resolution of those issues at a systemic level. The result would be suggested improvements to the way that judgments are currently delivered to mirror a digital world. The goal is to build a system where all suitable judicial and
tribunal decisions in New Zealand\textsuperscript{1} are made available through a single user-facing pipeline in a format that facilitates analysis by computer systems. This would help bring about a future where, through innovation in the market and the profession, access to the law and access justice is improved.

A report on our findings will be published in June-July 2020, and will be freely available to the public. Our goal is that this report will describe the current system and, if appropriate, make specific recommendations to bring about this future system.

With regards to the way in which judicial decisions are made available in New Zealand, we are seeking your input to help us understand the following:

1. How does the current system operate?
2. Why does it operate in that way?
3. What elements of the current system should be preserved and what can be improved?
4. How should that change occur, with emphasis on the needs of existing actors in the system?

Background to research topic

The task of conducting large scale analysis of judicial decisions is difficult, intensive, and time-consuming. Despite this, judgments represent a large potential legal dataset from which to draw conclusions and inferences about the operation of the justice system. However, there are systemic barriers to capitalising on this dataset: Firstly, gaining access to judgments can be a difficult process involving multiple agencies and, at times, some improvisation. Secondly, the format of those judgments is not amenable to being processed by computer systems.

Despite this, computer systems offer the potential to greatly overcome the aforementioned challenges of judicial decision analysis. They can make the process of finding, collating, and analysing judgments much easier, while helping us extract new kinds of insights across large numbers of judicial decisions.

Computational analysis of judgments can thereby make legal services easier to provide and to obtain, enhance access to justice and the law. This hope this will be immensely useful to policymakers, lawyers, judges, and citizens. We are interested in better understanding how the system could enable innovation on the potential of judgments as data.

\textsuperscript{1} We use “judgment” and “decision” interchangeably to refer to written decisions of Courts and Tribunals that would ordinarily be made public, and not those subject to suppression orders or other justified reasons for non-disclosure.
About the Project

The goal of the research project is to build a software prototype to capture the potential of judgments as data. The prototype would enable legal researchers to conduct large-scale analysis of judicial decisions in order to draw conclusions about the operation of the justice system. In conjunction with other forms of data and research analysis, this will better inform policy makers and leaders in the legal community about the access to justice needs faced by Court users and the kinds of issues being brought to the Court for resolution.

It is a proof of concept that will enable OpenLaw NZ to demonstrate the capabilities of software with a view to expanding it to use in other areas.

The project has been funded by the New Zealand Law Foundation through its Information Law and Policy Project. The research project is a collaboration between two separate groups:

- A team of independent legal researchers. They previously used judgments as data in order to identify access to justice barriers faced in the Accident Compensation jurisdiction.
- OpenLaw NZ, a registered charity devoted to access to legal information. OpenLaw is building open-source software that analyses the text of judicial decisions to extract useful information.

It is the team of researchers who are organising this discussion, with advice from OpenLaw NZ as an interested stakeholder with relevant expertise.

The project is intended to benefit the following groups and they will need to be included in relevant discussions at an appropriate time:

- the Judiciary
- the public
- Court users, including self represented litigants and those who are incarcerated
- legal practitioners (both lay and admitted)
- academic researchers
- policymakers and policy researchers
- independent researchers
- Court registry staff and other justice workers
- community legal services providers

---

2 Tom Barraclough, Warren Forster and Curtis Barnes worked together on various New Zealand Law Foundation funded projects analysing access to justice for ACC claimants using the Accident Compensation Act 2001’s dispute resolution processes. Tom and Warren continue to work with ACC in various ways on enhancing its dispute resolution processes.

3 William Parry, an experienced software programmer focussed on community led access to justice projects; and Andrew Easterbrook, a practicing Barrister and Solicitor of the High Court of New Zealand.
Current issues faced by the legal research project

(1) Access to judgments

We are interested to understand whether there is any coordinated system for making judgments available publicly. How do people get access to judgments in New Zealand? We are conscious of the Judicial Decisions Online service and the important work of the New Zealand Legal Information Institute. But how do these services interact and what support do they receive? These services can be contrasted with large commercial publishing databases. What provision exists for avoiding cost barriers, given these are primary legal materials? Could there be cost savings in the justice system by making similar information more freely available?

In the case of this research project, access has been generously supported by the Accident Compensation Corporation. ACC has provided us with copies of judgments that it holds in its databases. We are grateful for this assistance, but other researchers may not be able to benefit from such arrangements.

All of these databases may, to some degree, face issues of completeness. Why? Could this be avoided entirely?

(2) The format of those judgments

Judgments tend to be available in PDF and RTF formats. As a datasource, RTF formats can be too malleable. They do not preserve formatting. They can be altered by anyone with a word processor. As a datasource, PDF formats are not malleable enough, although can still be altered using commercially available software. In order to recover the text from PDFs at scale, users must often pay for proprietary optical character recognition (OCR) software in order to extract the text from them.

1. What are the benefits of PDF or RTF format?
2. What formats might be preferable given the overall requirements of the judgment delivery system?

Intended outcome

Our ambitions are high. Our vision is to see that all decisions of Courts and Tribunals in New Zealand will be made available in formats that facilitate their use as data in computational law contexts.

It is difficult to say what the outcome will be once decisions are available in this form: that is the point of doing it. The potential for innovation is significant. One use case, for example, is that there could be automated linking between New Zealand’s statute law and its case law, enhancing access to legal materials for the public, including small business and non-profit institutions.
One particularly exciting implication of access to this data is the way that it could facilitate new kinds of research by undergraduate students on legal technology applications. Some courses given in the United States of America, for example, give course credit for projects that use computers to achieve legal tasks, which tend to have a community justice focus, and often lead to legal tech spin-offs.

In the same way that legislation is available and reliable online through legislation.govt.nz, case law will be too.

**Key principles of this discussion**

Key aims or guiding principles of this proposal include:

- Open access to primary legal materials for the public and others.
- Promotion of rule of law values including access to the law.
- Encouraging innovation in dealing with the access to justice gap.
- Reducing the cost of legal services to the public.
- A research method for gaining insight into the impact of access to justice reforms and developments in the Court system over time\(^4\), including the increased use of ADR to divert disputes from the Court system.
- Benefit to public agencies responsible for assessing performance of the Court system.
- Accountability in the administration of justice through better access to data.
- Public access to publicly funded information.
- Availability of precedents and increased user engagement with primary legal sources, including associated development of secondary sources.
- Opportunities for scholarship and research at the undergraduate level.
- Respect for the independence and non-politicisation of the judiciary.
- Cost savings and enhanced competition for databases of legal decisions.
- Increasing New Zealand’s independence from large technology platforms and protection of open source open access initiatives.
- Accessibility of judgments (e.g. by the use of screen reader technology for people with visual impairments).

There is an established but growing field referred to as “computational law”. This refers to two things:

1. the simulation of legal processes and legal systems, including modelling legislation in computer code and emulating judicial reasoning processes; and
2. the more commonplace use of **computers to achieve legal tasks** such as document review in discovery, legal research platforms, contractual drafting and review, etc.

---

\(^4\) As noted by Justice Miller in a recent panel discussion, very few disputes occur now around sale and purchase of real estate because of the legal certainty created by the Auckland District Law Society standard form agreement.
This project belongs in that **second** category. It is important to emphasise that this project is NOT intended to encourage any of the following:

- Automation of judicial decision-making, judicial reasoning, or replacement of judges with computer systems.
- Tools to predict the outcomes of litigation in order to give legal advice.
- Putting the interests and requirements of computer systems ahead of litigants and other participants in the legal system, or compromising the Court’s ability to do justice in the case before it.
- Preferential access for OpenLaw NZ, the legal research team, or Brainbox Ltd.
- Under-cutting or replacing the New Zealand Legal Information Institute (“NZLII”).

**Key questions for engagement**

1. What can be done to enable all New Zealanders to have access by default to every published judicial decision in a machine-readable format?
2. Have we overlooked anyone who may be able to shed light on the answers to these questions?
3. How are judgments distributed now?
4. Why are they distributed this way?
5. How would it be better to distribute them? What do we mean by “better”?
6. What barriers are there to achieving this?
7. What are the potential risks if we are successful?
8. What are the potential opportunities if we are successful?

**Comparison data**

We are seeking data that might allow us to estimate the cost savings created by increasing access to judicial decisions in machine readable form.

**Key stakeholders**

We have identified key individuals or groups of stakeholders in relation to this proposal.

- JDO (judicial decisions online)
- Government Centre for Dispute Resolution
- The Judiciary and Chief Justice Helen Winkelmann
- The New Zealand Law Foundation
- Justice Miller as a member of the Advisory Review Committee for ILAPP.
- Dr Bridget Toy-Cronin and colleagues at the University of Otago Legal Issues Centre.
- NZLII and AustLII.
- Ministry of Justice.
- Ministry for Courts.
Conclusion

Thank you for your attention to this document, which has been produced by Tom Barraclough, Curtis Barnes and Warren Forster with funding from the New Zealand Law Foundation.

The research team can be contacted via Tom Barraclough by emailing tom@brainbox.institute or calling +6421900848.