Linked Democracy: Artificial Intelligence for Democratic Innovation

Marta Poblet, Pompeu Casanovas and Enric Plaza (eds.)

Proceedings

IJCAI 2017 Workshop

Melbourne, Australia

19 August 2017

Linked Democracy: Artificial Intelligence for Democratic Innovation

Proceedings of the IJCAI 2017 Workshop on Linked Democracy: Artificial Intelligence for Democratic Innovation

Marta Poblet, Pompeu Casanovas and Enric Plaza (eds.)

Melbourne, Australia
19 August 2017

With the support of:

RMIT University
La Trobe University
UAB Institute of Law and Technology
Artificial Intelligence Research Institute (IIIA-CSIC)

Cover image: Adapted from 'The Acropolis as viewed from the Mouseion Hill' by Christophe Meneboeuf (Creative Commons Attribution-Share Alike 3.0 Unported)

Foreword

The Workshop on 'Linked Democracy: Artificial Intelligence for Democratic Innovation' is one of the official workshops of the International Joint Conference on Artificial Intelligence (IJCAI 2017) held in Melbourne (19-26 August 2017). The goal of this workshop is to provide a multidisciplinary forum to address questions such as: How to model the interactions between people, data, and digital tools that create new spaces and forms of civic action in the digital era? How to analyse emerging properties and types of knowledge in these contexts? How to design socio-technical systems that effectively leverage data and knowledge for deliberation (or other types of participation) and collective decision making? Can we design the meta-rules of the emergent ecosystems?

The Workshop brings together participants from universities and research centers in Australia, New Zealand, Spain, Brazil, Israel, UK, and the USA. The workshop received 10 submissions, covering a number of different areas in AI (e.g. multi-agent systems and machine learning), economics, political sciences, and law. All submitted versions were reviewed by at least two members of the Program Committee. These proceedings finally include nine of these papers and an invited keynote speech by Patrick Keyzer.[1]

We sincerely thank the Program Committee members for reviewing all submitted papers and providing feedback to improve their revised versions. We are also grateful to the IJCAI 2017 chairs (Program Chair Carles Sierra, Workshops Chair Daniele Magazzeni and Local Arrangements co-Chair Andy Song) for their support in preparing this workshop. Last but not least, we would like to thank the participants who submitted their papers and afterwards produced the revised versions that are now composing these proceedings.

Marta Poblet, Pompeu Casanovas, and Enric Plaza

Workshop Chairs

[1] The submission not published in this volume can be found at Cohensius, G., Mannor, S., Meir, R., Meirom, E., & Orda, A. (2017). Proxy Voting for Better Outcomes. In Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems: 858-866.

Program Committee

Tanja Aitamurto (Stanford University, US)
Michal Araszkiewitz (University of Cracow, Poland)
Amir Aryani (Australian National Data Service, AU)
Thomas Bruce (Cornell University, US)
Danièle Bourcier (CNRS, France)
Alissa Centivany (University of Western Ontario, Canada)
Joel Chan (Carnegie Mellon University, US)
Yosem Companys (University of California Santa Barbara, US)
Mathieu D'Aquin (Open University, UK)
Louis de Koker (La Trobe University, AU)
Virginia Dignum (Delft University, The Netherlands)
Aldo Gangemi (LIPN-CNRS-Paris13-Sorbonne, France)
Asunción Gómez-Pérez (UPM, Spain)
Jorge González Conejero (UAB, Spain)
Guido Governatori (Data61, CSIRO, AU)
Juho Kim (KAIST, Korea)
Sabrina Kirrane (Vienna University of Economics and Business, Austria)
Mark Klein (MIT, US)
Gilly Leshed (Cornell University, US)
Karen Levy (Cornell University, US)
Wolfgang Mayer (University of South Australia, AU)
Brian McInnis (Cornell University, US)
Tarik Nesh-Nash (Mundiapolis University, Morocco)
Pablo Noriega (IIIA-CSIC, Spain)
Danuta Mendelson (Deakin University, AU)
Julian Padget (University of Bath, UK)
Ugo Pagallo (University of Turin, Italy)
Monica Palmirani (University of Bologna, Italy)
Jeremy Pitt (Imperial College, UK)
Iddo Porat (College of Law and Business, Israel)
Jason Potts (RMIT University, AU)
Victor Rodriguez Doncel (UPM, Spain)
Magda Roszczynska-Kurasinska (University of Warsaw, Poland)
Giovanni Sartor (European University Institute, Italy)
Geoffrey Stokes (RMIT University, AU)
Markus Stumpfner (University of South Australia, AU)
Tom van Engers (University of Amsterdam, The Netherlands)
Bart Verheij (University of Groningen, The Netherlands)
Roland Vogl (Stanford University, US)
Julian Waters-Lynch (RMIT University, AU)
Mark Whiting (Carnegie Mellon University, US)
Adam Wyner (University of Aberdeen, UK)
John Zeleznikow (Victoria University, AU)

Table of Contents

Open Rights or Secret Risk Assessments? New Challenges for Public Law in an Age of Artificial Intelligence and the Law

Keynote Speech

Patrick Keyzer[1]

[1] La Trobe University, Bundoora, VIC 3086, Australia
p.keyzer@latrobe.edu.au

1 The Cautionary Tale of "Robo-Debt"

Centrelink, Australia's welfare agency, recently contacted the Commonwealth Scientific and Industrial Research Organisation, Australia's leading public scientific agency, and asked its "Data61" team to review its electronic data-matching system with the Australian Taxation Office, the "Online Compliance Intervention". The purpose of the Online Compliance Intervention was to data-match tax records and welfare payment records, work out whether there were any discrepancies, and then correct them (Commonwealth Ombudsman, 2017). It makes sense for governments to ensure that welfare benefits are paid to people who are genuinely in need, and to ensure that benefits are not paid to people who have paid employment. Properly set up and operated, such a system could help ensure that resources are most effectively deployed to prevent poverty and minimise welfare fraud.

However this is not how things worked in practice, in a great many cases. Instead, from about midway through last year and well into this year, Centrelink's data-matching system was conflating annual earnings figures from the Tax Office with fortnightly income profiles used by Centrelink to assess welfare payment needs. It transpires that the "Matching Techniques" Protocol deployed an algorithm to calculate totals for a number of financial years, but since Centrelink doesn't return detailed fortnightly data about a given *year's* earnings, the algorithm simply averaged the amounts over the relevant periods. This approach was prone to errors, particularly in circumstances where a beneficiary might have made transitions in and out of work (increasingly common in a casualised labour market).

Regrettably, the errors produced by this defective approach were compounded by the administrative approach taken to the recoupment of the debts. Letters were sent out automatically, without any internal, human review of the relevant calculations. These identified debts of hundreds or even thousands of dollars and then required the recipients to contact Centrelink within 21 days to explain the discrepancy, or pay the debt under penalty of enforcement. The letters did not include contact telephone numbers for the compliance team, so people seeking assistance contacted Centrelink through a general call line that resulted in long waiting times. It transpires that the call line was staffed by people who had little knowledge of the Online Compliance Intervention because they had not been trained.

The characterisation of these letters has been a bone of contention. They were described by the departmental secretary in a subsequent Senate Inquiry as “clarification letters” – as providing *opportunities* for welfare beneficiaries to “clarify” the discrepancy between the identified debt and the beneficiary’s personal records. However to ensure that this “clarification” was provided within a particular time period, the Government authorised private debt collectors to follow-up the letters with telephone calls in return for a 10% bonus payment. The “clarification” was thus outsourced by the government department to the “customer”, and to be conducted under the shadow of a penalty.

After many complaints, the Commonwealth Ombudsman initiated an investigation. The Ombudsman’s report concluded that the enforcement regime imposed unreasonable burdens on welfare beneficiaries and Centrelink staff. The Ombudsman said that while it was reasonable for Centrelink to ask beneficiaries for assistance in explaining discrepancies in its records, that the 21-day timeframe was unreasonable, and that the success of the Initiative depended on its usability. Usability in turn depended on the accuracy and completeness of departmental information, which was questionable in many cases. (Peter Hanks QC has questioned this, noting that the Ombudsman did not address the question whether the Social Security Act can create a debt presumptively, and whether the Department of Human Services could shift the fact-finding task to the individual: Hanks, 2017). The Ombudsman also found that the requirement that people keep records over six or seven years was unreasonable, in part because beneficiaries had not been forewarned of this requirement. The Report also outlined problems related to planning and implementation, lack of consultation, and a failure to plan for or properly mitigate risks. In addition, the Ombudsman found that Centrelink’s assistance and customer support was defective (it has been reported that there were 42 million unanswered phone calls in a month) and that staff had not been adequately trained to support customers and to deal with complaints.

The Robo-Debt Controversy could be dismissed as a classic case of “garbage in, garbage out” – the “algorithm” was flawed, and if it was fixed, then, according to the Ombudsman, it was otherwise fair for Centrelink to request “clarification”. Perhaps the reference to Data 61 means that the Government will take expert advice to ensure that the design of such systems will be improved in the future. But the Robo-Debt Controversy is a cautionary tale. With the increasing use of artificial intelligence systems in public administration, where is the place for human oversight and procedural justice? What steps should we take to protect the most vulnerable people in our society from the public service machines? This paper raises some questions. Answering them will take more work.

***

The deployment of artificial intelligence systems makes a lot of sense in large-scale, routine work that requires no or only minimal and manageable discretion (Perry and Smith, 2014). For example, a program developed by the United States Department of Veterans Affairs manages disability claims and has completely replaced human public servants by requiring applicants to fill in a detailed questionnaire that is then processed by software. Properly designed and implemented, these systems can speed up the operation of administrative justice and enhance transparency. Questions arise, of course, when artificial intelligence initiatives disrupt—legally-sanctioned bureaucratic authority. Notwithstanding the risks, it is likely that governments will continue to seek out new ways to automate such systems, in order to save revenue, particularly in environments of austerity (Perry and Smith, 2014) (let alone artificially-produced scarcity).

In a recent lecture series at Penn State, Justice Cuellar of the California Supreme Court, a former Professor of Law and Information Technology at Stanford Law School, has identified a number of potential side effects from automated public decision-making. Cybersecurity risks are an obvious example. However the impact of automation on *dialogue* is by far the most important:

> Implicit in democratic governance is an aspiration for dialogue and exchange of reasons that are capable of being understood, accepted, or rejected by policymakers, representatives of organized interests, and members of the public.
>
> Except when computerized decisions can rely on relatively straightforward, rule-like structures, difficulties will arise in supplying explanations of how decisions were made that could be sufficiently understood by policymakers and the public.

Cuellar (2016) also remarked:

> This is not to say that the status quo is any deliberative panacea. On the contrary, it is easy to criticize the current administrative state for its lack of opportunities to allow the public to participate in decisions. Yet the growing reliance on automated computer programs to make sensitive decisions in the administrative state will only complicate what little deliberation does occur.

Do the principles of Australian law provide adequate normative resources for dealing with the challenges ahead? This is a question that could inspire many academic papers. In this paper I will focus on just one, but an important one: What happens when we don't know why the machine has made the decision it has made?

2 Risk Assessment on Secret Grounds

In 'The Minority Report' (1956), the science fiction writer Philip K Dick famously invented 'Precrime', a government agency (later popularized by Tom Cruise in a very ordinary movie) which enabled the surveillance and apprehension of people who would commit murders in the future. Today, suspected terrorists can be detained without charge on suspicion of future harm and sex offenders can be sent to prison on the basis of a risk assessment in circumstances where they have committed no new crime (McSherry and Keyzer, 2009). The United Nations Human Rights Committee has said in several decisions that imprisoning a person on the basis of a risk assessment in the absence of a fresh crime and criminal trial is arbitrary detention and incompatible with the International Covenant on Civil and Political Rights (Keyzer, 2011) but this continues to be done, as Australia does not honour its international human rights obligations (O'Donovan and Keyzer, 2014). In the absence of human rights norms, Australia is left with weak constitutional protections and the common law. Can these principles ensure procedural justice in pre-crime scenarios?

The use of secret algorithms in risk assessments is a new, worrying development in the criminal justice system. Recently, a Wisconsin trial court sentenced a man named Eric Loomis to six years' imprisonment for participating in a drive-by shooting (Liptak, 2017). In sentencing, the trial court considered a report derived from a software product called Compas, produced by a company called Northpointe Inc. Compas uses an algorithm to weigh a number of risk factors and produce an actuarial risk assessment of a person accused of a crime. At trial, the prosecutor submitted a Compas report about Loomis that found him to be a high risk of violence, a high risk of recidivism, and a high pretrial risk. Loomis sought details about the software algorithm so that he could challenge its conclusions. Northpointe declined to release any details about how its Compas algorithm calculates risk on the basis that it is proprietary, and commercial-in-confidence. Loomis appealed the ruling of the trial judge.

The process of actuarial risk assessment is well explained by Brad Johnson in his paper "Prophecy With Numbers" (2006), in terms worth setting out at some length:

> Psychiatrists and psychologists have employed a number of methods for determining risk with respect to human behaviour, which include … clinical assessment, actuarial risk assessment and actuarially informed clinical assessment which combines elements from each. The difference between clinical and actuarial assessment is reflected in the type of data relied on in order to determine the level of risk—clinical assessment relying primarily on data about the person being assessed and actuarial assessment relying on data from a population of individuals who share a number of attributes in common with the person being assessed, thus allowing statistical comparative judgements. …
>
> Actuarial risk assessment departs from clinical assessment methods by examining populations of released offenders in order to identify attributes that are associated with an increased risk of recidivism. The data with respect to recidivism rates collected from multiple sample populations of released offenders can be used to make some simple inferences. The relative frequency of recidivism for a particular sample may be used to make a probability statement about the chance of an individual, who shares the attributes that define the population, committing a future offence. Alternatively the relative frequencies for various populations may be compared to determine which samples display a higher level recidivism, which in turn is believed to indicate a greater risk of recidivism. The process of establishing relative frequencies with respect to recidivism begins by examining an initial population of released offenders for a specific period of time which yields relative frequencies for those who reoffend and those who do not. The sample population being investigated also allows researchers to look for attributes that are associated with recidivism. The initial population can

> then be analysed by specifying further attributes that break the population down into more clearly defined demographic groups in the hope of identifying greater recidivism rates for specific populations.

Elsewhere, Bernadette McSherry and I have written that the use of risk assessment scales may be justifiable for the purpose of treatment in clinical environments (2009), but problems with these scales are amplified by their use in legal forums, where they can be ‘prone to manipulation and misinterpretation’ (Sullivan, Mullen and Pathé, 2005, p 319), leading to unnecessary detention due to false positive findings that the individual concerned is at risk of harming others. Importantly, all of these scales are based on variables that are derived from analyzing groups, giving rise to the ‘statistical truism that the mean of a distribution tells us about everyone, yet no one’ (Cooke and Michie, 2010). Ian Coyle and Robert Halon (2013) have further observed that:

> The law guarantees that a decision will be made but it does not guarantee outcomes. Yet that is precisely what the law seeks to require of those engaging in the task of risk-analysis of dangerousness. It is time, once and for all, to acknowledge that estimates derived from actuarial tests cannot predict the future behavior of individuals with anything approaching that implicit in the legal minimum standards of proof.

Returning to the Loomis case, the trial court used a COMPAS report to justify incarceration. The court said that “You’re identified, through the COMPAS assessment, as an individual who is at high risk to the community. In terms of weighing the various factors, I’m ruling out probation because of the seriousness of the crime and because your history, your history on supervision, and the risk assessment tools that have been utilized, suggest that your [sic] extremely high risk to re-offend.” Loomis’ counsel led evidence from an expert witness, Dr Thompson, who outlined the pitfalls of relying on actuarial risk assessment in a sentencing context. The State of Wisconsin did not offer any witnesses to counteract this evidence, instead arguing that the court’s conclusion did not rely on the COMPAS report and, if it did, any reliance was a “harmless error”.

One of the appeal points was whether the trial court’s use of COMPAS at sentencing violated Loomis’ constitutional right to due process because Loomis could not challenge the scientific validity of the assessment due to Northpointe’s proprietary claim over the software algorithm. Loomis argued that it was unknown which criminogenic factors COMPAS utilizes, and how it weighs them. He relied on *Gardner v Florida* (430 U.S. 349, 351 (1977)). In Gardner, the defendant was convicted of first degree murder and the trial court sentenced him to death. The defendant appealed because the trial court deemed certain portions of the pre-sentence investigation report to be confidential and refused to disclose the information to counsel. The US Supreme Court held that the defendant was denied due process because the trial court had imposed a sentence “at least in part, on the basis of information which (Gardner) had no opportunity to deny or explain”. The Court of Appeals accepted Loomis’ submission that the same principle applied here, and concluded that the “apparent limited ability of (the) defendants to investigate the tool" unfairly prevented them from assessing its scientific validity.

Wisconsin appealed to the US Supreme Court. In his appellate submissions, Loomis argued that:

> the only basis for COMPAS are the ipse dixit statements from Northpointe that it does what it says; that although we do not know how it weighs the criminogenic factors, we should just take the risk assessment as true. To do so, however, violates Mr. Loomis' (and other defendant's) right to due process because information upon which the trial court is relying for sentencing is secret and confidential. As the Court of Appeals noted, there is a lack of transparency.

Wisconsin, for its part, defended COMPAS (citing Brennan, 2009). Remarkably, Wisconsin argued that "Loomis claims the COMPAS report may have been inaccurate, but he cannot prove it because he does not know how COMPAS calculates risk". Quite. Instead, Wisconsin argued that Loomis knew what questions the COMPAS evaluation asked and he knew the answers to the questions – and on *that* basis he could contest the answer to specific questions on the COMPAS evaluation if he thought that the correct answer was different than the answer entered, and *that* procedure satisfied the constitutional due process requirement. Specifically, according to the State of Wisconsin, "Due process does not require disclosure of the formulas used to determine risk".

Remarkably, the US Supreme Court rejected the appeal. The Court was likely influenced by an amicus curiae brief filed, on the Court's request, by the US Solicitor-General. The Solicitor-General opined that given "the highly limited purpose for which petitioner's ability to counter the factual information on which the assessment relied, the Wisconsin Supreme Court correctly declined to find a due process violation. But that is not to say that the use of actuarial risk assessments at sentencing will always be constitutionally sound." While the issues the petition raised were conceded to be important, the Solicitor-General said that any "constitutional error in considering (the) petitioner's COMPAS score was likely (to have been) harmless".

What a remarkable occasion to apply the maxim *de minimis non curat lex*.

The Wisconsin Supreme Court has since published a guideline judgment relating to COMPAS which has rather confusingly said that while COMPAS Reports cannot be determinative, they nevertheless may be regarded as *relevant* in sentencing. *How* relevant will, it seems, remain a mystery. Perhaps tacitly acknowledging the weakness of this reasoning, the Wisconsin court has imposed several prophylactic guidelines: first, any "presentence investigation report ("PSI") containing a COMPAS risk assessment filed with the court must contain a written advisement listing" its limitations, and second, if used in sentencing, the following cautions need to be applied:

- "The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined.

- Because COMPAS risk assessment scores are based on group data, they are able to identify group of high-risk offenders-not a particular high-risk individual.
- Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism.
- A COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations."

On this basis, "*if used properly* with awareness of the limitations and cautions, a circuit court's consideration of a COMPAS risk assessment at sentencing does not violate a defendant's right to due process" (emphasis added).

Not long after the Supreme Court delivered its judgment rejecting the Loomis appeal, Professor Shirley Ann Jackson, President of the Rensselaer Polytechnic Institute in New York, asked the Chief Justice of the United States, the Hon John Roberts Jr., "[c]an you foresee a day when smart machines, driven with artificial intelligences, will assist with courtroom fact-finding or, more controversially even, judicial decision-making?" Roberts CJ replied, "It's a day that's here, and it's putting a significant strain on how the judiciary goes about doing things."

It seems remarkable that the constitutional right to due process would not protect the defendant in a criminal trial, and ensure that person's access to information used against them. Would Australian common law principles of procedural fairness operate to protect a person placed in a similar position in Australia?

3 "Preventive Exile" after Failing the Character Test

I'm not aware of any Australian case where a court had to consider the procedural justice implications of the use of secret algorithms. However the Australian Government is, apparently, actively considering the development of risk assessment tools, and it is conceivable that similar issues might arise. A Cabinet briefing note leaked in the first half of 2016 proposed the introduction of "a *visa risk assessment tool* that establishes an intelligence-led threat identification and *risk profiling capability* incorporating immigration as well as national security and *criminality risk* for visa applicants".[1]

Ian Coyle and I have written elsewhere (2016) about the use of character testing in the immigration system (and the next few paragraphs rely heavily on that paper). Specifically, in late 2014 the Minister for Immigration and Border Protection issued Direction 65 to supplement section 501 of the *Migration Act* 1958 (Cth), which enables the Minister or a delegate to cancel a visa held by a noncitizen convicted of an offence on the basis that they have failed a 'character test'. A person is presumed to

[1] David Lipson, 'Leaked Government document outlines tougher migration program, increased monitoring of refugees', ABC News, 4 February 2016.

fail the character test if they have a ‘substantial criminal record’, defined as a criminal conviction attracting a sentence (or cumulative convictions and sentences, adding up to) of 12 months or more.

The removal of people pursuant to this regime takes place without prior notice being given (presumably to prevent absconding) and is often effected in the early hours of the morning by armed personnel.[2] Once arrested and removed from their homes and families, these people are typically taken to an immigration detention centre such as Christmas Island. Christmas Island is a lovely name for an island but it is very remote – some thousands of kilometres from Australia in the middle of the Indian Ocean. It is a long way from family, friends and supports. Here, detainees may wait many months or even years for their case to be heard, assuming they are able to secure legal assistance to do so. This plainly raises significant concerns about their ability to access the justice system to challenge their removal.

With Ian Coyle, I have explored the use of a ‘risk assessment’ as a basis for decision-making about removals – and considered whether it is compliant with proper forensic standards. We have argued that if risk assessments are to be undertaken, they need to be undertaken properly, and with a nuanced appreciation of the limitations of forensic tools. Serious questions can be raised about the utility of actuarial risk assessment tools. However the risk remains that governments will devise regulations that remove the power of litigants and courts to question these tools. Applying knowledge of group tendencies to individual offenders within actuarial risk assessment approaches can have dire consequences when transferred to court settings in high stakes cases where liberty or citizenship is at stake (McSherry and Keyzer, 2009). Problems identified with the *use* of actuarial-based scales in relation to individual offenders were acknowledged in the guideline judgment delivered after *Loomis v Wisconsin.* But it is the gloss on *Gardner v Florida* that is significant in an age of algorithmic governance. It is difficult to prevent an involuntary shaking of the head when a company’s proprietary interests are elevated above the right to liberty.

In China, a social credit system has been devised to rate people on their social and financial behavior. It is said that this new system will “allow the trustworthy to roam everywhere under heaven while making it hard for the discredited to take a single step.” (Hawkins, 2017). While horrifying, and so horrifying to be scarcely believable, surveillance is not new, and has been carried out for eons. The Chinese Communist *dang’an,* or secret personal file, tracks a citizen’s information from their high school grades, to their behavior at university, to their perceived political sympathies in adult life. The file can affect a person’s career prospects and pension entitlements. The Tibetan writer Tsering Woeser has described the dang’an as “an invisible monster stalking you” (Jacobs, 2015).

There has been an appreciable rise in the development and deployment of risk assessment tools to judge us all. We want to remove risk from our lives and we expect our governments to do this. In the commercial sphere, we are all affected by the administrative justice meted out by electronic platforms and services that we use for

[2] *Eden v Minister for Immigration and Border Protection* [2015] FCA 780.

transport, to do shopping, and to employ assistants. We buy into these regimes by rating people ourselves.

Actuaries say that we should work with all available information, and that it would be wrong not to. But where is the place for the presumption of innocence, let alone the possibility of rehabilitation, in the coming dystopia? Do we have the legal tools to challenge the risk assessments?

References

1. Australia. Commonwealth Ombudsman, Report No 2 of 2017.
2. Tim Brennan et. al., ‘Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System’, (2009) 36 *Criminal Justice and Behaviour* 21.
3. David Cooke and Christine Michie, ‘Limitations on Diagnostic Precision and Predictive Utility in the Individual Case: A Challenge for Forensic Practice’, (2010) 34 *Law and Human Behavior* 259.
4. Ian Coyle and Robert Halon, ‘Humpty Dumpty and Risk Assessment: A Reply to Slobogin’, in Patrick Keyzer, ed., *Preventive Detention: Asking the Fundamental Questions*, Intersentia, 2013.
5. Ian Coyle and Patrick Keyzer, ‘The removal of convicted noncitizens from Australia: is there only a ‘minimal and remote’ chance of getting it right?’ (2016) 41(2) *Alternative Law Journal* 86
6. *Gardner v Florida* 430 U.S. 349 (1977).
7. Peter Hanks QC, ‘Administrative law and welfare rights: a 40-year story from Green v Daniels to “robot debt recovery”’, Australian Institute of Administrative Law Conference, Canberra, July 20, 2017.
8. Amy Hawkins, ‘Chinese Citizens Wants the Government To Rank Them’, Foreign Policy, May 24, 2017.
9. Andrew Jacobs, ‘A Rare Look Into One’s Life On File in China’, Sinosphere, 15 March, 2015.
10. Patrick Keyzer, ‘The International Human Rights Parameters for the Preventive Detention of Serious Sex Offenders’, in Bernadette McSherry and Patrick Keyzer, eds., *Dangerous People: Policy, Prediction and Practice*, Routledge, 2011.
11. Bernadette McSherry and Patrick Keyzer, *Sex Offenders and Preventive Detention*, The Federation Press, 2009.
12. Melissa Perry and Alexander Smith, ‘iDecide: the Legal Implications of Automated Decision-Making’, University of Cambridge, Cambridge Centre of Public Law Conference, 2014
13. Danny Sullivan, Paul Mullen and Michele Pathe, ‘Legislation in Victoria on Sexual Offenders: Issues for Health Professionals’, (2005) 183(6) *Medical Journal of Australia* 318
14. Mariano-Florentino Cuellar, ‘Artificial Intelligence and the Administrative State’, The Regulatory Review: A Publication of the Penn Program on Regulation, December 19, 2016.

Towards a Linked Information Architecture for Integrated Law Enforcement

Wolfgang Mayer[1], Markus Stumptner[1], Pompeu Casanovas[2,3], and Louis de Koker[2]

[1] University of South Australia, Adelaide, Australia
[2] La Trobe Law School, La Trobe University, Melbourne, Australia
[3] UAB Institute of Law and Technology, Universitat Autònoma de Barcelona, Spain

Abstract. Law enforcement agencies are facing an ever-increasing flood of data to be acquired, stored, assessed and used. Automation and advanced data analysis capabilities are required to supersede traditional manual work processes and legacy information silos by automatically acquiring information from a range of sources, analyzing it in the context of on-going investigations, and linking it to other pieces of knowledge pertaining to the investigation. This paper outlines a modular architecture for management of linked data in the law enforcement domain and discusses legal and policy issues related to workflows and information sharing in this context.

Keywords: law enforcement, investigation management, linked data.

1 Introduction

Investigations conducted by law enforcement agencies (LEAs) are increasingly reliant on effective collection and analysis of information that may be obtained from a variety of sources, internal and external to the organization [1]. Investigations generally follow an iterative process of information collection, assessment, investigation planning, execution, and brief of evidence preparation where each step either produces new information or relies on information collected earlier in the process.

Information collected within the organization include information about individuals, organizations, objects and entities of interest, witness statements, evidence obtained from crime scenes, communications intercepts and the results of forensic analysis. This information may be complemented and integrated with data such as financial transactions, travel and immigration records, and criminal history that are obtained from external sources. In addition, documentation about the investigation process and data provenance must be maintained in order to establish that evidence submitted to court had been obtained within the law and policies relevant to the investigation.

Accessing data as well as linking and integrating them in a correct and consistent way is a pressing challenge in particular when underlying data structures and access methods change over time. Lack of interoperability between information systems within and across organizations remains one of the prevalent concerns of investigators [2]. Investigations are delayed by poor information management practices that result in information being unavailable or not being available in a timely manner, poor information quality, and cumbersome manual approval and information retrieval procedures.

The project Integrated Law Enforcement (ILE), conducted by the Data to Decisions Cooperative Research Centre (D2D CRC)[1], aims to develop a platform where investigators can manage the information collection, analysis, and processes pertaining to a case through a consistent single user-facing platform. The project has been developing technological solutions for information management, linking, and analysis that are tailored to the needs of investigators. An extensible software architecture for searching, linking, and integration of data sources forms one of the corner stones of the project. The platform will eventually include analytic services that can be invoked by investigators. The data management architecture is complemented with a state of the art user-facing portal and an analysis of legal aspects pertaining to workflows and information sharing.

Effective linking, integration, and analysis of data requires breaking down data "silos" and opening up legacy systems within organizations to make information accessible, establishing procedures and technical infrastructure to effectively and timely share information across organizational boundaries, creating data standards to facilitate interpretation and analysis of the body of collected data, and automating, where possible, analysis and semantic enrichment of data [3].

Data integration in this context raises serious legal compliance and good governance challenges. Compliance with existing laws and principles is a pre-condition of the whole process [4]. Transparency and privacy should be preserved to foster trust between citizens and national security and law enforcement agencies. A 2015 literature review on online data mining technology intended for law enforcement broadly singled out eight main problems (crimes, investigative requirements) in 2015 [1]. Separately, some criminologists warned against the profound effect of automated data collection on the traditional criminal justice system, as it could undercut the due process safeguards built into the traditional criminal justice model [5]. It is our contention that this technological modelling should be performed under the protections of the rule of law.

In this paper, we present the overall system architecture for information sharing and outline the related legal issues pertaining to workflows and information exchange in the context of policing investigations. Our work has resulted in a data access framework for law enforcement which provides a comprehensive data and meta-data model including provenance, security, confidence, links and timeline information related to entities and links. This meta-data layer spans a Knowledge Graph-like view [7] of information pertaining to entities relevant to investigations. The resulting data and meta-data model serves as the foundation for information use, governance, data quality protocols, analytic pipelines and exploration of search results.

[1] http://www.d2dcrc.com.au/

2 Information Sharing in Law Enforcement

Timely information sharing domain is crucial for the success of many investigations in the law enforcement domain. Unfortunately, many investigations are stalled by one or more of a number of impediments related to effectively sharing information among investigators and organizations [2]. In the following we highlight a selection of issues relevant in context of linked information access.

Among the *technical impediments*, internal information silos and cumbersome information access procedures are common. Investigators routinely enter the same queries across a multitude of legacy information systems and manually collate and integrate the results. Lack of information access mechanisms for investigators in the field hamper the timely acquisition of information in electronic form and information may not be updated in timely manner. In absence of automated alerts, investigators may be unaware that new information relevant to a case has become available unless they manually issue queries periodically or rely on informal personal connections to receive notifications. As a result, relevant information may be missed even though it had been available in an information system. Data quality varies greatly as data quality standards are often not enforced and instead left to the individual user.

Workflows and policies may impact upon investigations. Where approvals for actions are required, for example expenditure approval for call records requests, antiquated policies and work processes may still rely on paper forms and manual approval which result in excessive delays, in particular if approvals are sought outside of normal office hours. Here, automation and electronic means of requesting and obtaining warrants and approvals would streamline the investigation process.

Legal issues relate to restrictions on information use and sharing. For example, information obtained under a warrant for a specific investigation may not generally be used in the context of other investigations. Similarly, agencies are generally subject to restrictions on what information they can share with other agencies [2]. Even where information sharing may be legally permitted, many organizations, concerned about the implications of breaching the law, are prone to adopt prudential attitudes and policies that perhaps may unnecessarily restrict what can be shared. *Information security* and access control are challenging issues when multiple systems and organizations are involved. It is challenging to guarantee comprehensive and secure access to a large number of users accessing a multitude of information systems across organizational boundaries. Moreover, there is interaction between analytics and security attributes as new information derived from automated analytic processes must be classified using appropriate security policies to avoid inadvertently disclosing otherwise inaccessible information. Determining appropriate classification and access restrictions can be challenging in organizations.

3 System Architecture

An open architecture for data/meta-data management and analytic processes has been defined. It translates the best practices from Enterprise Application Integration to the

“Big Data” analytic pipelines [6]. Our work addresses aspects related to data and meta-data modelling and storage, modelling and execution of analytic processes, and efficient execution of analytic processes across multiple analytic tools and data sources. Central to this architecture is a method for effective semi-interactive entity linking and querying of linked data (akin to automatically generated linked ontologies such as YAGO [13]). The project intends to realize a comprehensive data management framework that relies on a well-defined share data and meta-data model supported by vendor-agnostic interfaces for data access and execution of processes comprising analytic services offered by different tools.

The overall architecture of the ILE platform is shown in Figure 1. A federated architectural model has been adopted, where one or more instances of the ILE platform can be deployed and access a number of external data sources. Each instance may provide query and analytic services to the front-end applications and can obtain data from other instances and external sources on demand. This approach is necessary as data in external sources is usually controlled by external organizations and may change at any time. Moreover, organizational policies in this context rarely support traditional Extract-Transform-Load ingestion processes across organizational boundaries.

The ILE platform provides *programmatic interfaces (APIs)* to front-end applications to access data and invoke analytic services. The interfaces expose the platform’s services using a uniform data format and communication protocol. The APIs can be accessed from a desktop front-end where investigators can search and enter information as well as invoke services. Mobile applications for investigators may be developed in future versions of the platform.

Each instance maintains a *Curated Linked Data Store*, that is, a set of databases that collectively implement a knowledge-graph like structure comprising entities and their links and meta-data. This curated data store holds facts and meta-data about entities and their links whose veracity has been confirmed. This data store is used to infer the results for queries and to synthesize requests to external sources and other instances if further information is required. As such, the linked data store implements a directory of entities and links enriched with appropriate meta-data and source information such that detailed information can be obtained from authoritative sources that may be external to the system. This approach is needed as data in the law enforcement domain is dispersed among a number of systems owned and operated by different agencies. As such no centrally controlled database can feasibly be put in place in the foreseeable future.

The information contained in the linked data store is governed by an *Ontology* that defines the entity types, link types, and associated meta-data that is available among the collective platform. The ontology acts as a reference for knowledge management/organization and aids in the integration of information stemming from external sources, where it acts as a reference for linking and translating information into a form suitable for the knowledge hub. The ontology has been designed specifically for the law enforcement domain and includes detailed provenance information and meta-data related to information access restrictions. It is explicitly represented and can be queried. All information within the ILE platform is represented in the ontology in order to facilitate entity linking and analysis.

The ILE ontology is too large to reproduce it in full in this paper; it comprises 19 high-level domain concepts which are further refined into a total of ~140 concepts and a taxonomy of ~400 specialized relationship types. It has been documented in [8]. The ontology conceptualizes the domain on three levels: meta-level where concept types are captured, the type level, where domain concepts are represented in terms of types, and the instance level, where instance-level data is represented and linked. For example, the meta-level defines EntityType, RelationshipType, and MetaAttributeType. Their instances on the level below represent persons, organizations, (and more broadly a hierarchy of object types), concrete domain relationships that may be established between objects (for example that a Person works for an Organization), and meta-data attributes related to access control, provenance, and temporal validity.

These domain concepts are closely aligned with the draft National Police Information Model (NPIM), complemented with relevant aspects drawn from the NIEM standard[2] and concepts related to case management. The provenance model is an extension of PROV-O [9]. The instances of the domain concepts form the objects comprising the Knowledge Graph on the lowest layer in the ontology. The aforementioned concepts are complemented with classes and objects representing data sources linked to the domain information stored therein as well as schema mapping information required to translate between the external source and the ontology model adopted within the federated architecture.

This multi-level modelling method has been adopted to provide a modular and extensible knowledge representation architecture. The semantic technologies that underpin our platform facilitate incremental addition of elements to the ontology, and phasing out of obsolete concepts can be implemented via meta-data annotations interpreted by the underlying information systems. Changes in information representation received from external parties can be addressed by ontology matching techniques and machine learning methods for information extraction and linking. Profound changes in the information acquisition pipelines however would require changes to the underlying information system. Our modular architecture has been designed to accommodate such changes.

Information from external sources is sought based on a *catalog* of data sources that are available to the system, each with a corresponding *adapter* that communicates with the external systems and rewrites the information and meta-data into the ontology used within the ILE platform [11]. Our platform spans several sources, including an entity database (Person, Objects, Location, Event, and Relations), a case management system, and a repository of unstructured documents.

Information received from external systems is passed through an *ingestion and enrichment* pipeline where entities are extracted [14], enriched with meta-data (provenance and access restrictions) and linked to the knowledge graph in the linked data store.

[2] https://www.niem.gov/

Analytic services include entity extraction from unstructured text [14], entity linking, similarity calculation and ranking. Services provided by commercial tools, such as network analysis and entity liking/resolution solutions, can be integrated in the modular architecture.

Automation services will provide workflow orchestration and alert notices if new information relevant to a case becomes available. Workflow services will facilitate the enactment of work processes such as acquiring authorization and warrants. The automation services component is pending implementation.

Cross-cutting technical concerns, including access control and user management, logging, monitoring and other deployment facilities, have been omitted in this architecture view. Our implementation builds on open source big data technologies (Hadoop/Spark, polyglot persistence, message queues, and RESTful interfaces). The technical building blocks are outlined in [8].

Fig. 1 draws the overall architecture and plot the direction of legal workflow processing.

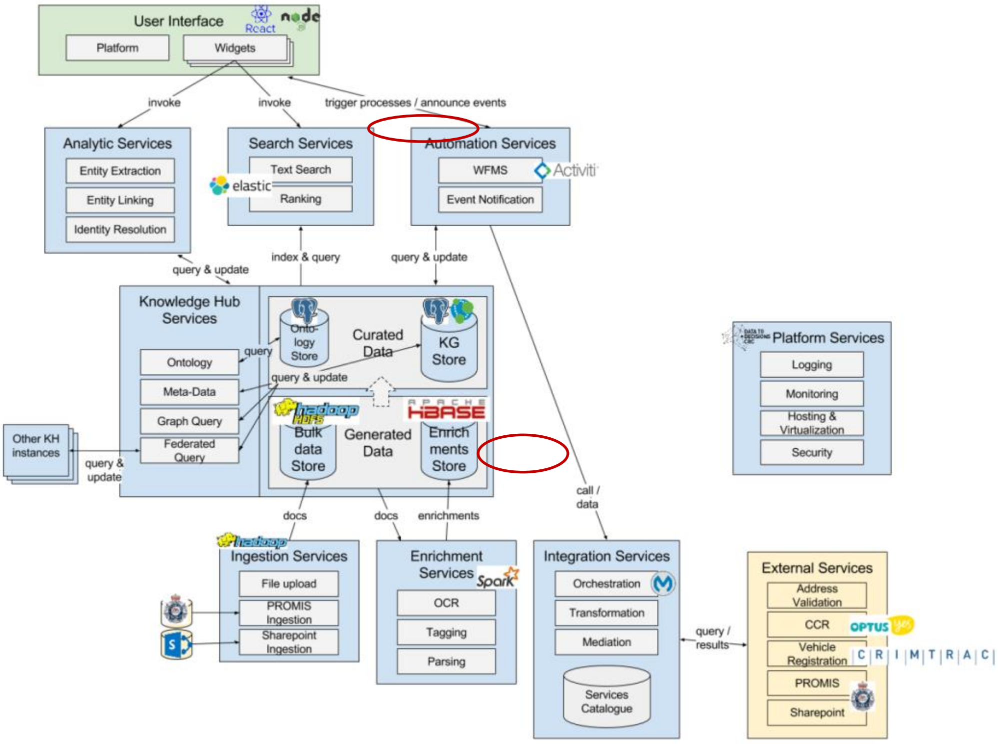


Fig. 1 Architecture overview and legal workflow processing

4 Legal and governance issues

A key concern is the incorporation of legal risk management and compliance constraints into the workflow execution to ensure observance of and compliance with the applicable legal rules, for example agency and privacy rules as well as internal agency policies and procedures. Natural Language parsing can be used to elicit event specifications that could then be translated to business rules in an executable formal language and issued to an event processor in the knowledge hub [16]. These rules would be used to check and guarantee conformance of analytic processes/workflows and data usage. [16] provides support for extracting data from a variety of sources (relational databases, CSV files, JSON, and XML), for modeling it according to a vocabulary of the user's choice and for integrating multiple data sources. This process deserves a closer attention, because (i) it implies LEAs cooperation, and (ii) must be compliant with Australian law.[3]

The D2D CRC's law and policy team outlined for discussion a set of high level principles that may guide the development of an appropriate framework: (i) engender public confidence in government use of data and analytic tools, (ii) develop principles for data governance in National Security Law Enforcement (NSLE) agencies; (iii) employ clear and consistent principles in developing legal frameworks, (iv) improve processes to enhance effective use of data within NSLE agencies, (v) ensure the continued effectiveness of the oversight regime as technologies and NSLE agency practices evolve, (v) disentangle elements of technological change associated with 'Big Data', (vi) maintain data integrity and security in a high volume environment, (vii) ensure fair and appropriate use of data analytics, (viii) use appropriate systems for data matching, data integration or federated access that takes account of benefits and risks; (ix) ensure efficient, appropriate, and regulated sharing of specific data for NSLE purposes [17].

In a recent survey we carried out on the state of the art of Compliance by Design (CbD) [12], we found that the passage from Business to Legal CbD mainly follows a semantic path, in which Natural Language Processing (NLP), non-monotonic defeasible logic and inferential reasoning are combined with enriched annotated legal sources (e.g. described according linked data standards). This is aligned with recent developments in e-business[4] [26] and e-government [27]. Architectures are deemed

[3] We consider primarily investigations conducted by Australian law enforcement agencies, where compliance with Australian laws governing these investigations and subsequent legal proceedings is paramount.

[4] ISO/IEC 42010:2007 defines "architecture" as: "The fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution". It has been fleshed out by [26], and [27] [28] for the e-government architecture. See esp. ADM Architecture Requirements Management, and the Architecture Compliance steps defined at TOGAF 9.1, Part VII: Architecture Capability Framework Architecture Compliance. http://pubs.opengroup.org/architecture/togaf9-doc/arch/

to be understandable, robust, complete, consistent and stable. [28] has proposed a comprehensive approach to develop interoperable European e-government services adapting and extending the existing enterprise architecture requirements. The investigation shows that at least half —i.e. not all— of the 30 requirements identified are adequately addressed by enterprise architectures (EA).[5] It concludes with ten interoperability challenges that should be taken into account and addressed when providing pan-European e-government services (PEGS) across Member State borders. Quoting at length: (i) critical success factors should be identified, (ii) an EA framework for PEGS should be built upon widely accepted principles and strategies, (iii) it should comprise architecture design principles and guidelines to reason about alternative design strategies, (iv) in order to facilitate stakeholder management, it should refer to abstract stakeholder classes and roles in interoperability projects and determine drivers for their engagement, (v) the creation of contents can be improved through a methodology that supports the capturing of requirements from business-driven needs, policy implementation processes and other strategic aspects in order to establish common path and to increase the acceptance of architecture outputs among stakeholders (vi) another methodology should describe how to define interoperability specifications on semantic and organizational level, which can be used as a basis for collaboration agreements, (vii) a detailed design of each architecture should identify relevant model fragments and should be based on a commonly agreed architecture description language, (viii) there are missing guidelines and methods that describe how to transition and to govern architectures in multi-stakeholder environments, (ix) several independent implementations of PEGS have to be coordinated, extended and sustained over time (e.g. it should integrate appropriate assessment methodologies to measure specifications and the compliance of solutions with the underlying collaboration agreements, (x) other assessment methodologies can help to determine the level of business standardizations in a domain and to appraise the maturity of market solutions in order to detect appropriate ways forward.

This is a valuable programme. Likewise, we have also devised one close to it with the Australian framework in mind. However, as [28] underline as well, business languages do not completely match all governance and security requirements. Interoperability frameworks do not enable an anticipatory management [29].

Legal compliance is complex, even in relation to national laws where the jurisdiction concerned is a unified, non-federal national state. There are several methodologies and languages to represent norms using formal rules —e.g. Regorous and LegalRuleML [31]—, but there are not *fully* automated ways to carry out such a task. Legal norms must be interpreted in particular fields according to the specific domains to which they apply, anticipating the possible risks and unintended side effects. In addition, ethical principles can nuance or mould this interpretation according to different jurisdictions — e.g. Fair Information Practices in USA, or Data Protection Principles

[5] The 30 requirements obtained in the survey of the literature have been structured into six categories [28]: project management (PM), stakeholder management (ST), service development (SD), interoperability layers and architecture viewpoints (LV), building blocks (BB), and collaboration agreements (CA).

similar to the brand new General Data Protection Regulation in Europe. Similarly, information governance rules and policies differ between private corporations and state agencies.

At a more general level, legal scholars have noticed that the protection of relative civil rights such as privacy does not necessarily entail tradeoffs [21][6]. Nevertheless, as we have already suggested, there are many ways to comply with rule of law requirements, depending on the plurality of legal constraints and constitutional specifications. Apparently, protections for civil rights are not as clear —and arguably as strong— in Australia as in the EU, where the police and their criminal intelligence functions operate subject to well-developed data protection and privacy norms. In contrast to a more comprehensive, integrated EU approach, it could be argued that public transparency and operational secrecy are, for example, not as finely balanced under current Australian law [19]. Contrary to European provisions, the *2017 Australian Productivity Commission Inquiry Report on Data Availability and Use*, excludes national security data [20]. As asserted by the Report, governments use data to monitor and investigate compliance and implement enforcement actions. They retrieve, extract and analyse information from publicly available sources (Open Source Intelligence, OSINT) in a way that can also be regulated [22].

Having a closer look, problems about fragmentation and interoperability are analogous in both Australia and Europe. Different as they might be, the post-facto investigations about the Abdelsam brothers in the Bataclan crisis in Paris [30] and the inquest into the deaths arising from the Lindt Café siege in Sydney[7] have come to similar conclusions. Cooperation among state departments and agencies; and between Law Enforcement Agencies (LEA), can and should be improved.

These conclusions are not limited solely to security issues but can be also extended to the coordination of public administration and the legal system in policy domains. For instance, in many situations, problems might arise "because of gaps of information flow between the family law system, the family violence system, and the child protection system: in many circumstances, important information is not being shared among courts and agencies and this is having a negative impact on victims, impeding the 'seamlessness' of the legal and service responses to the family violence".[8]

Disparity is produced as well across all Australian jurisdictions. At a federal level the *Privacy Act 1988*, for example, regulates the handling of personal information by the federal government and the private sector. The Act does not extend to state governments. Some states have their own comprehensive frameworks. In Victoria, for example, the Privacy and Data Protection Act 2014 (Vic) contains the following Information Privacy Principles (IPPs) which apply to all information held by the Victorian public sector (including the police and a contracted service provider): (i) Open and transparent management of personal information, (ii) Sensitive information, (iii)

[6] International human rights law distinguishes between absolute and relative rights. Absolute rights —such as freedom from slavery, torture and servitude— cannot be suspended, restricted, or limited for any reason. Non-absolute or relative rights are those which stand in the various private and legal relations, and can be discussed, re-defined or qualified.

[7] http://www.lindtinquest.justice.nsw.gov.au/

[8] Australian Law Reform Commission (2010), as quoted in [20].

Right to anonymity/pseudonymity, (iv) Notification of collection, (v) Purpose test for use / disclosure, (vi) Direct marketing restrictions, (vii) Cross border disclosure, (viii) Government-related or unique identifiers, (ix) Data quality, (x) Data security, (xi) Access and correction. South Australia, on the other hand, only has an administrative instruction requiring government agencies to comply with a set of Information Privacy Principles while Western Australia does not currently have a comprehensive legislative privacy regime.

Australia has a comprehensive oversight regime in relation to national security and law enforcement agencies. Different bodies have however oversight over different agencies or have oversight over closely-defined aspects of a range of agencies. The fragmented nature of the oversight framework in Australia "will be challenged by an environment where NSLE agencies collaborate more closely in a Big Data framework" [19]. But to reach this milestone, it is our contention that the information integration process that takes place on the platform through reusable ontologies and vocabularies requires a broader regulatory framework. To overcome the patchwork of disparate and sometimes contradictory legal constraints, we will work within an intermediate implementation level, setting what can be called an "anchoring institution" between the semantic tools of the platform and LEAs (end-users).

This set of intermediate conceptual rules constitute a semantic web regulatory model (SWRM), i.e. a specific cluster of guidelines to regulate the information flow, establishing a system of check and balances between LEA's investigative powers and their use of semantic technology [22]. This is an indirect strategy for Compliance by Design (CbD) purposes, in which police officers might set forth internal and external controls, and adopt a conceptual scheme to implement privacy and security principles including ethics as a main component, i.e. at the intermediate level of linked democracy [23]. To encompass both behavioural and informational trends, we use the expression 'Compliance through Design' [CtD] [12] . This means that the increasing of pressure on human compliance management resources in the security area can be taken into account [25]. The crucial point is the coexistence of both artificial and human decision-making and information processes.

Likewise, 'linked democracy' can be defined as "a meso-level approach to both online and offline innovations that elucidates the interactions between people, technology, and data in particular settings, providing a framework of analysis to understand the emerging properties (and tensions) of these interactions" [24]. Therefore, public principles such as transparency, accountability and security could be graduated and connected within particular investigations according to their weight at their specific implementation level. This entails the emergence of different notions, degrees and values of legal compliance, enhancing their semantic side, and outstripping the traditional obstacles of operating from separate information silos.

It is worth noticing that from this pragmatic approach, interoperability does not only mean 'semantic interoperability' —the creation of a common meaning for information ex- change across computational systems— but *systemic* interoperability. That is, the ability of complex systems to interact, share, and exchange information. It focuses onto the coordination of practices, including human behavior, organizational structures, tools, languages, and techniques [23]. Establishing such a model, translating legal and systemic conditions to institutional and computational constraints and requirements, is the next step.

Acknowledgements

This research was partially funded by the *Data to Decisions Cooperative Research Centre (D2D CRC)*, with participation of the Spanish Project *DER2016-78108-P*. Views expressed herein are however not necessarily representative of the views held by the funders.

References

1. Edwards, M., Rashid, A., Rayson, P. A systematic survey of online data mining technology intended for law enforcement. ACM Computing Surveys (CSUR), 48 (1) (2015): 1-56.
2. Scheepers, R., Whelan, C., Nielsen, I., Burcher, M., Integrated Law Enforcement Project, Qualitative End User Evaluation, Baseline Report. Technical Report, Data 2 Decisions CRC, 2017.
3. Mayer, W., Stumptner, M., Grossmann, G., Jordan A. Semantic Interoperability in the Oil and Gas Industry: A Challenging Testbed for Semantic Technologies. In AAAI Fall Symposium on Semantics for Big Data, Arlington, Virginia, November 2013.
4. Law and Policy Program. Big Data Technology and National Security - Comparative International Perspectives on Strategy, Policy and Law: Australia (Data to Decisions CRC), 2016.
5. Marks, A., Bowling, B. and Keenan, C. Automatic Justice? Technology, Crime and Social Control. In: R. Brownsword, E. Scotford and K.Yeung (eds), The Oxford Handbook of the Law and Regulation of Technology, Oxford University Press, pp. 705-730, 2017. Available at SSRN: https://ssrn.com/abstract=2676154 [2015]
6. Data to Decisions CRC Big Data Reference Architecture, vol. 1-4, Technical Report, Data to Decisions CRC, 2016.
7. Heath, T., Bizer, C. Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool Publishers, 2011.
8. Grossmann, G., Kashefi, A.K., Feng, Z., Li, W., Kwashie, S., Liu, J., Mayer, W., Stumptner, M. Integrated Law Enforcement Platform Federated Data Model, Technical Report, Data 2 Decision CRC, 2017.
9. Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S. and Zhao, J. Prov-o: The PROV ontology. W3C recommendation, 2013.
10. Stumptner, M., Mayer, W., Grossmann, G., Jiu, J., Li, W., De Koker, L., Mendelson, D., Bainbridge, B., Watts, D., Casanovas, P. An Architecture for Establishing Legal Semantic Workflows in the Context of Integrated Law Enforcement, Workshop on Legal Knowledge and the Semantic Web (LK&SW-2016), International Conference on Knowledge Engineering and Knowledge Management, Bologna, Italy, Nov. LNAI, 2017 (forthcoming).
11. Bellahsene, Z., Bonifati, A., Rahm, E. Schema Matching and Mapping. Data-Centric Systems and Applications, Berlin, Heidelberg: Springer, 2011. DOI: 10.1007/978-3-642-16518-4.
12. Casanovas, P; González-Conejero, J. Technical Bases for Compliance by Design (CbD). CRC D2D Deliverable, May 2017 (updated July 2017).

13. Suchanek, F.M., Kasneci, G. and Weikum, G. YAGO: a core of semantic knowledge unifying Wordnet and Wikipedia. In Proceedings of the 16th international conference on World Wide Web, pp. 697-706, ACM, May 2007.
14. Del Corro, L., Gemulla, R. Clausie: clause-based open information extraction. In Proceedings of the 22nd international conference on World Wide Web, pp. 355-366, ACM, May 2013.
15. Cardellino, C., et al. Licentia: A Tool for Supporting Users in Data Licensing on the Web of Data. Proceedings of the 2014 International Conference on Posters & Demonstrations Track-Volume 1272. CEUR-WS. org, 2014.
16. Gupta, S., Szekely, P., Knoblock, C., Aman, G., Taheriyan, M., Muslea, M. Karma: A system for mapping structured sources into the Semantic Web. In E. Simperl, P. Cimiano, A. Polleres, Ó. Corcho, and V. Presutti, editors, The Semantic Web: Research and Applications - 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012. Proceedings, LNCS 7295, pp. 430-434, Springer, 2012.
17. Law and Policy Program. Big Data Technology and National Security - Comparative International Perspectives on Strategy, Policy and Law: Australia (Data to Decisions CRC, 2016).
18. Casanovas, P., De Koker, L., Mendelson, and David Watts. "Regulation of Big Data: Perspectives on strategy, policy, law and privacy." Health and Technology (2017): 1-15.
19. Bennet-Moses, L., de Koker, L. Open Secrets: Balancing Operational Secrecy and Transparency in the Collection and Use of Data for National Security and Law Enforcement Agencies, CRC Report, 2017.
20. Australian Government. Data availability and use. Productivity Commission Inquiry Report, No. 82, 31 March 2017.
21. Pagallo, U. Online Security and the Protection of Civil Rights: A Legal Overview. Philosophy & Technology 26 (2013): 381–395.
22. Casanovas, P. Cyber Warfare and Organised Crime. A Regulatory Model and Meta-Model for Open Source Intelligence (OSINT). In R. Taddeo and L. Gkorioso, Ethics and Policies for Cyber Operations, pp. 139-167: Dordrecht: Springer International Publishing, 2017.
23. Casanovas, P., Mendelson, D. & Poblet, M. A Linked Democracy approach to regulate health data. Health and Technology (2017), DOI: 10.1007/s12553-017-0191-5
24. Poblet, M. and Plaza, E. Democracy Models and Civic Technologies: Tensions, Trilemmas, and Trade-offs, 2017, *arXiv preprint arXiv:1705.09015*.
25. Watts, D., Bridget Bainbridge, B., de Koker, L., Casanovas, P., Smythe, S. Project B.3 A Governance Framework for the National Criminal Intelligence System (NCIS), Data to Decisions Cooperative Research Centre, La Trobe University, 30 June 2017.
26. Open Group Standard TOGAF Version 9.1 Document Number: G116. ISBN: 9789087536794.
27. Brous, P., Janssen, M., Vilminko-Heikkinen, R. Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles. In H. J. Scholl et al. International Conference on Electronic Government and the Information Systems Perspective EGOVIS 2016, LNCS 9820, pp. 115-125, Springer International Publishing, 2016.
28. Mondorf, A., Wimmer, M.A. Requirements for an Architecture Framework for Pan-European E-Government Services. In H. J. Scholl et al. International Conference on Electronic Government and the Information Systems Perspective EGOVIS 2016, LNCS 9820, pp. 135-150, Springer International Publishing, 2016.

29. Mondorf, A. and Wimmer, M., Contextual Components of an Enterprise Architecture Framework for Pan-European eGovernment Services. In *Proceedings of the 50th Hawaii International Conference on System Sciences*, HICSS, 2017, pp. 2933-2942.
30. Bureš, O., 2016. Intelligence sharing and the fight against terrorism in the EU: lessons learned from Europol. European View, 15(1): 57-66.
31. Sadiq, S., Governatori. G. A methodological framework for aligning business processes and regulatory compliance. In: Jan van Brocke and Michael Rosemann, editors, Handbook of Business Process Management, pp. 159-176, Springer, 2010.

Bounded-Monitor Placement in Normative Environments

Guilherme Krzisch †, Nir Oren ‡, and Felipe Meneguzzi †

† Pontifical Catholic University of Rio Grande do Sul, Brazil
‡ University of Aberdeen, Scotland, UK
guilherme.krzisch@acad.pucrs.br, n.oren@abdn.ac.uk, felipe.meneguzzi@pucrs.br

Abstract. In order to sanction non-compliant agents, norm violations must be detected, which in turn requires norm monitoring. This paper examines the problem of monitor placement within a normative multi-agent system under budgetary constraints. More specifically we consider a system containing (1) a set of possible monitors able to determine the state of a subset of the domain; (2) costs associated with deploying the monitors; and (3) a set of norms for which compliance must be monitored, and which, if violated, result in a penalty. We seek to identify which combination of monitors maximizes the system's utility. We formalize the problem and evaluate approximate solutions using several heuristics, empirically demonstrating their efficiency.

1 Introduction

Since agents within open multi-agent systems cannot be assumed to share goals, obtaining desirable outcomes requires a coordination mechanism, and *norms* [5] are often used to perform this coordination. While regimented norms – which prevent an agent by design from undertaking undesirable behavior – are widely often assumed, a substantial body of work has shown the advantages of using approaches based on enforcement [8, 10]. These latter approaches, which allow norms to be violated, require a mechanism that monitors for norm compliance or violation and applies sanctions when appropriate [11]. In turn, *norm enforcement* can be performed either by some organization or by other autonomous agents in the system [14].

Norm enforcement assumes that violation and compliance can always be detected [6, 7, 11]. Such an assumption is, however, clearly unrealistic. First, in a large system, the cost of monitoring is very high [15]. Second, there are often portions of the environment that are not fully observable, and which monitors cannot access. There is therefore a need to investigate how norms should be monitored.

Previous work on the monitoring of norms has considered how norms should be modified so as to be monitorable [2], and whether related states can be observed which may indicate upcoming norm violations [1]. In this paper, we consider instead how monitors should be deployed within a system, assuming that such deployments have a cost, so as to monitor the most important norms. Our approach builds on ideas from the planning literature, and in Section 2, we introduce the necessary concepts from this domain and formalize the notion of a norm. Section 3 introduces monitors and formally describes the problem we are addressing. We describe several approaches to addressing the problem in Section 4. Finally, Section 5 evaluates our solutions, and we conclude by considering related work (Section 6) and directions for future research (Section 7).

2 Background

In this section we introduce concepts from the planning literature, used to formalise the system and our solution. Following this, we describe norms within our system.

2.1 Planning

We build on classical planning, which assumes finite, fully observable and deterministic systems, and adapt the definitions from Ghallab *et al.* [9, Ch. 2]. Systems that follow classical planning semantics assume that actions in the domain cause transitions between states, and are specified in terms of sets of predicates.

Definition 1 (Predicate and State). *A predicate in a first-order language $\mathcal{L}$ is composed of a symbol and zero or more terms. Each term can be either a constant or a variable; a predicate is ground if it does not contain variables. We denote as $|\mathcal{L}|$ the number of ground predicates in this language. A state is a set of ground literals (i.e., positive or negative predicates) in $\mathcal{L}$.*

We use the operator $\models$ to specify that a state (a set of predicates) satisfies a logical formula, i.e., that a formula is a model of the state.

Classical planning makes a closed-world assumption — that if a state does not specify a predicate, then this predicate does not hold in that state. We write $s \models P$ where P is a set of ground predicates, if the conjunction of these ground predicates is satisfied by s. We formalize action execution, and its effects on the environment as follows.

Definition 2 (Planning Operator and Actions). *A planning operator is represented by a triple $o = \langle name(o), pre(o), eff(o)\rangle$, where $name(o)$ is the description of o. $pre(o)$ and $\mathit{eff}(o)$ are set of predicates representing the planning operator's preconditions and effects. An action a is a ground instance of a planning operator, and is* applicable *in state s only if $s \models pre(a)$. The result of applying action a to state s is a new state s', such that $s' = (s/\mathit{eff}^-(a)) \cup \mathit{eff}^+(a)$, where $\mathit{eff} = \mathit{eff}^+ \cup \mathit{eff}^-$ (such that $\mathit{eff}^+ \cap \mathit{eff}^- = \emptyset$) and eff^- is the set of negated predicates and eff^+ is the set of positive predicates.*

We specify the dynamics of our multiagent system in terms of a transition function following classical planning semantics in Definition 3, and the initial states of the system in terms of planning problem instances, as per the following definitions.

Definition 3 (Planning Domain). *If $\mathcal{L}$ is a first-order language with finite sets of predicates and constants, a planning domain in $\mathcal{L}$ is a state-transition system $\Sigma = \langle S, A, \gamma\rangle$, where $S \subseteq 2^{|\mathcal{L}|}$ is a subset of all possible states; A is the set of all ground instances of planning operators; $\gamma(s, a)$ is a state-transition function defined as follows: if $a \in A$ and a is applicable to $s \in S$, it returns the next state $s' \in S$, which is the result of applying action a to state s.*

Note that S is closed under γ, i.e., given a state $s \in S$, all states reachable from applying action a in s are also in S.

Definition 4 (Planning Problem). *A planning problem is defined as a triple $\mathcal{P} = \langle \Sigma, s_0, g \rangle$, where Σ is the state-transition system, $s_0 \in S$ is the initial state of the problem and g, the goal, is a set of ground predicates.*

2.2 Norms

In open and dynamic societies, self-interested agents cannot be assumed to share the same set of goals. In this context, norms can be used to regulate and coordinate behavior [13, Ch. 14]. For our work, we adapt the definition of a norm from [12]. We consider two norm types: obligations and prohibitions; the former specifies behavior that must be followed by agents, while the latter specifies behavior that must be avoided.

Definition 5 (Norm). *A norm is a tuple $n = \langle \mu, \chi, \rho, C \rangle$, where:*

- $\mu \in \{obligation, prohibition\}$ *represents the norm's modality;*
- χ *is a set of ground predicates that represents the* context *to which a norm applies, i.e., a norm is applicable in state s if $s \models \chi$;*
- $\rho \in A$ *represents the* object *of the norm's modality;*
- C *is the cost or penalty* to the society *which occurs if the norm is violated.*

Example 1. The following norm requires an agent to drive on the left side of the road if they are in England; a violation causes harm to the society worth 20 units of utility.

$$n_0 = \langle obligation, \mathit{at}(\mathit{England}), \mathit{driveLeft}(a, b), 20 \rangle$$

We now describe when a norm is considered to be violated by an agent.

Definition 6 (Norm Violation). *A norm $n = \langle \mu, \chi, \rho, C \rangle$ is violated in state s by an agent a iff:*

- *$s \models \chi$; and*
- *agent a either: executes action ρ in state s and $\mu = prohibition$; or does not execute action ρ in state s and $\mu = obligation$.*

A violated norm has an undesirable impact on the society, as encoded by its cost C. To dissuade agents from violating norms, when such a violation is detected, an enforcer applies a penalty to the agent. In this work, we do not consider the nature of this penalty, assuming instead that it is sufficiently large to prevent the agent from violating a norm *if such a violation can be detected*. We must therefore consider how to place monitors so that violations are detected, while minimizing monitor costs. We refer to this as the *bounded-monitor placement problem*, and describe it in more detail in the next section.

3 Bounded-Monitor Placement Problem

We consider a set of monitors able to determine whether some combination of predicates is, or is not satisfied. Formally, we define a monitor as follows.

Definition 7 (Monitor). *A monitor $m = \langle P, D\rangle$ consists of a set of predicates P, and a* deployment cost $D \in \mathbb{R}$. *We refer to the predicates of a monitor m as P_m, and to its cost as D_m.*

A set of monitors can be used to monitor more complex combinations of predicates. Some monitors can conceivably detect the status of a predicate related to multiple norms, or when combined, can be used to determine the status of a norm that individual monitors cannot. We formalize the combination of monitors aiming to cover sets of norms as a *Monitor Placement*.

Definition 8 (Monitor Placement). *A monitor placement MP is a tuple $\langle M_1, M_2\rangle$, where M_1 and M_2 are sets of monitors, representing the capability of observing predicates in the current and in the next state, assuming an action was performed.*

A monitor placement detects *a norm violation $\langle \mu, \chi, \rho, C\rangle$ iff given a state s and s' such that s' is the result of applying an action ρ at s; and $s \models \chi$, one of the following holds:*

1. *$\mu = prohibition$ and $s \models \bigwedge\{P_{m_1} | m_1 \in M_1\}$ and $s' \models \bigwedge\{P_{m_2} | m_2 \in M_2\}$.*
2. *$\mu = obligation$ and $s \models \bigwedge\{P_{m_1} | m_1 \in M_1\}$ and $s' \not\models \bigwedge\{P_{m_2} | m_2 \in M_2\}$.*

*The cost **C** of a monitor placement is $\sum_{m \in M_1 \vee m \in M_2} D_m$, while the utility **U** is $\sum_{n \in N} C$, where N is the set of norms detected by this placement.*

By introducing the concept of an available budget, we define our problem of placing monitors in a system as follows.

Definition 9 (Bounded-Monitor Placement Problem). *A bounded-monitor placement problem is encoded as a triple $\langle M, N, B\rangle$ where M is a set of monitors, N is a set of norms, and $B \in \mathbb{R}^+$ is a budget.*

A solution to the problem is a monitor placement MP such that its cost is smaller than, or equal to, the budget.

The set of possible solutions for a monitor placement problem is exponential in the worst case, and in the next section, we suggest several approaches to finding solutions.

4 Solution

4.1 Brute-force

A brute-force approach is a trivial solution to this problem. It considers all possible combinations of available monitors and returns the best one. We can clearly see that this is impractical for large problems, as its complexity increases exponentially given the size of the possible monitors set input: specifically, it has a time complexity of $O(2^{2|M|})$. We use this approach as a baseline against which we compare the remaining heuristics.

4.2 Mapping from norms to monitors

The main drawback of the brute-force approach is that it includes a large number of irrelevant solutions while enumerating all possible solutions. We can reduce this overhead by computing a mapping from norms to monitor placements, i.e., by finding the set of all monitor placements capable of monitoring each norm.

With this mapping, we can compute possible solutions by choosing one monitor placement for each norm. Generalizing, we have $\prod_{i=1}^{|N|}(|\mathbf{MP}_{n_i}|+1)$ possible solutions, where $\mathbf{MP}_{n_i}$ is the set of monitor placements able to monitor norm n_i; since we also need to consider whether it is practical to monitor a given norm (e.g., when there is no available budget to do so), we need to add the empty monitor placement set. In the best case we have only one possible monitor placement for each norm, and in the worst case we have all possible combinations of available monitors for MP_{M_1} and MP_{M_2}, for each norm. Therefore, the number of solutions ranges from $2^{|N|}$ to $4^{|M||N|}$.

Given this exponential complexity, it is clear that both the brute force and monitor mapping approaches cannot scale up to larger problem sets. We must therefore consider heuristics for addressing the problem, which we describe next.

4.3 Naive Approximate Solution

To improve performance compared to the brute-force approach we introduce a simple approximate solution whose purpose is to serve as a a baseline for comparing the accuracy of the other solutions. This solution iterates over monitors ranked by their expected probability of detecting norm violations. To rank monitors, we consider the number of norms that a single monitor can partially detect — a monitor can partially detect a norm if it has at least one predicate of the norm's context or of the preconditions of the norm's action ρ. The intuition here is that choosing monitors that can partially observe several norms leads to a final monitor placement that can detect many existing norms.

This approach, however, does not capture essential parts of the problem. First, it does not consider the norm's penalty and monitor's cost. Second, it has an overly strong assumption that joining monitors that can partially detect norms will lead to a monitor placement that can actually detect norms. We can therefore enhance this approach by using the mapping describe in Section 4.2 together with a greedy search.

4.4 Greedy Solution

We propose two approaches with different heuristics using the mapping structure introduced in Section 4.2. By using a heuristic to rank the best monitor placements, we can avoid searching through an exponential solution search space. More specifically, we select the best monitor placement candidate for each norm. The resulting heuristic sacrifices optimality for efficiency, running in linear time.

Our base algorithm (used for both of our heuristics) is described in Algorithm 1. It starts by creating a mapping between norms and monitor placements, as described in Section 4.2. After this, it adds monitors to an initially empty monitor placement $currentMP$, while budget is available. Within each iteration, it selects one norm, gets a monitor placement able to monitor this norm, and adds the already selected monitors

Algorithm 1 Greedy algorithm

```
1: procedure FINDAPPROXIMATESOLUTION(N:Norms)
2:     build mapping from norms to monitor placements
3:     $currentMP \leftarrow \{\}$
4:     while hasBudget do
5:         $n \leftarrow extractMaxNorm(N)$
6:         $mp \leftarrow getMinMP(n)$
7:         $currentMP \leftarrow currentMP \cup mp$
       return $currentMP$
```

($currentMP$) to the monitor placement. We now describe two heuristics to speed up this algorithm.

4.4.1 Norm Independence Heuristic Our first heuristic considers norms to be monitored completely independently of each other when choosing monitors in order to substantially prune the search space of the problem. We first need to define which norm *extractMaxNorm(N)* in line 5 of Algorithm 1 chooses. Here, we select the norm with the highest penalty, in order to increase the value of **U** (the sum of each individual norm penalty), i.e., $extractMaxNorm(N) = \arg\max_{n \in N}\ C_n$.

The other decision required is which monitor placement is selected by the *getMinMP(n)* method (line 6). For this, we select the placement with the lowest cost, as it is a good monitor placement able to monitor norm n. Thus, $getMinMP(n) = \arg\min_{MP \in \mathbf{MP}_n}\ \mathbf{C}_{MP}$.

This approach does not consider the intersection of monitors that are able to monitor multiple norms; it selects monitor placements independently of the others. Consequently, we improve this solution in what follows by introducing the concept of *current cost* to try to find a better approximate solution for our problem.

4.4.2 Add and Update Heuristic The structure of this heuristic is similar to the previous approach. However, instead of using the cost of each monitor placement as the metric to select the one with lowest budget, we use its *current cost*. The current cost of a given monitor placement MP is computed as per Formula 1 below, and considers only the cost of monitors that were not already selected in a previous iteration (and thus not in the set of monitors of *currentMP*).

$$\mathbf{CC}_{MP} = \sum_{m \in M_{MP} \wedge m \notin M_{currentMP}} D_m \tag{1}$$

$$getMinMP(n) = \underset{MP \in \mathbf{MP}_n}{\arg\min}\ \mathbf{CC}_{MP} \tag{2}$$

$$extractMaxNorm(N) = \underset{n \in N}{\arg\max}\ \frac{C_n}{\mathbf{CC}_{getMinMP(n)}} \tag{3}$$

The *getMinMP(n)* method is implemented as per Formula 2, which we use to compute the next norm to be monitored using the *extractMaxNorm(N)* method, implemented in Formula 3. This heuristic chooses the norm with the highest value based on the ratio between its penalty and the lowest current cost of the set of monitor placements. By using the current cost, we disregard the costs of monitors that have already been chosen in past iterations, yielding better estimates. In the next section we empirically evaluate the different approaches proposed in this section.

5 Experiments and Results

To empirically evaluate our approaches, we automatically generate sets of norms and sets of monitors with increasing complexity. In our experiments, we assumed the set of monitors M_2 of a monitor placement can in effect see all actions that were executed from the states visible by M_1; i.e., we assume that our monitor placement is always able to check the next state after applying a given action. While this is a strong assumption and makes our problem easier to solve, it still captures the exponential nature of the bounded-monitor placement problem. Our experiments are based on standard planning problems [], and our main domain is the *drink-driving* domain, where agents are able to drive between cities, and there is a norm stating that it is forbidden to drive while drunk; we also tested with *blocksworld, depots, dwr, easy_ipc_grid, gripper, logistics* and *robby* domains.

There are two metrics to consider when analyzing results: time performance and accuracy. When comparing time performance we use the brute-force approach as a baseline for the approximate approaches. In Figure 1 we can see that the brute-force approach becomes intractable for a relative small number of norms, while the approximate approaches remain linear as the number of norms increases.

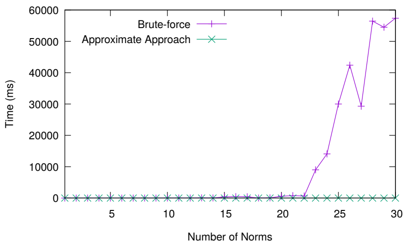


Fig. 1. Time Efficiency of Brute-Force and Approximate approaches, with timeout of 60 seconds

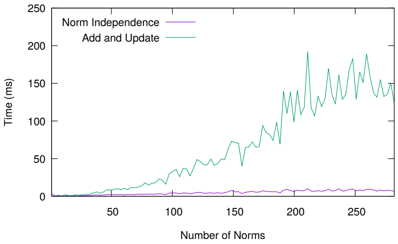


Fig. 2. Time Efficiency of both greedy approaches, smoothed using a sample of 100 data points and interpolated using splines

Figure 2 shows the time performance of both greedy solutions. The *Add and Update Solution* is worse in this metric than the *Norm Independence Solution* because it needs to recompute the current cost at each iteration. It is still fast compared to the brute-force approach, being able to find a solution for a problem with almost 300 norms in under one second.

We perform two experiments to compare the accuracy of the results of the approximate approaches. First, in Figure 4, we show the relative accuracy compared with an optimal solution to the problem using the brute-force approach. Note that, as we are comparing with the brute-force approach, these results are limited to small problems that this approach can solve[1]. In this experiment, both greedy approaches outperform the naive solution; between the two greedy approaches, the second one (*Add and Update Heuristic*) has, in almost all domains, greater accuracy, and — in all cases — a smaller standard deviation. For the *depots*, *gripper* and *logistics* domain, the second solution achieves optimal accuracy; this can be explained as these domains become intractable even for small number of norms, and thus the number of problems in this experiment, for these domains, is also small. The increase in performance from the first to the second greedy approach is relatively small for these domains; while it is relatively large for the *drinkdriving* and *robby* domains.

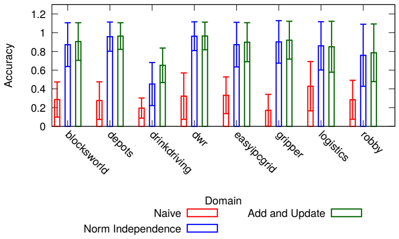


Fig. 3. Accuracy of approximate approaches, compared to the maximum U possible; error bars represent one standard deviation of uncertainty

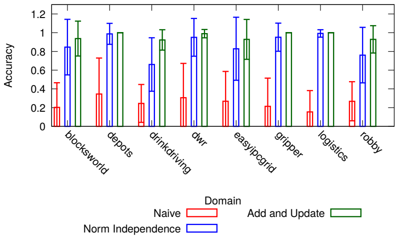


Fig. 4. Accuracy of approximate approaches, compared to a brute-force solution; error bars represent one standard deviation of uncertainty

To investigate if the relationships found for the first experiment hold for large problems, we perform additional experiments, shown in Figure 3. As these problems cannot be solved in a timely manner using a brute-force approach, in order to calculate their accuracy we compare them with a perfect solution that could monitor all norms, but that does not necessarily need to respect the available budget. This perfect solution has the maximum value of **U**, which can be unattainable for actual solutions to these problems; therefore, in this experiment we are interested in the relative accuracy between the approximate approaches, and not in how close they are to the perfect solution.

We can see the same pattern in this experiment; the greedy approaches perform better than the naive solution, with a slight advantage for the *Add and Update Heuristic*. The increase in performance from the first to the second greedy approach remains large for the *drinkdriving* domain, while for other domains it is relatively small.

From the experiments we conclude that brute-force solution is intractable for all but small problems, while approximate approaches can solve large problems. The accuracy of the greedy approaches is better than the naive solution, with a slight advantage to the second greedy approach. This advantage is more noticeable for small problems; for large problems — as we do not have the value for the optimal solution — this difference is smaller.

[1] We set a timeout of 30 seconds for this experiment.

6 Related Work

Other authors also dropped the assumption that a monitor has full observability in the system. Criado's [4] approach is similar to our work; their norms are more expressive, and for their solution they use the CEF algorithm which uses a greedy approach. However, they do not perform an empirical evaluation, or consider whether their approach is able to actually monitor any norm, as there are no guarantees proven for their proposed algorithm. Alechina *et al.* [2] models a set of queries that a monitor can ask in a state, i.e., a monitor may not be able to distinguish between two states. They then modify the set of norms to a new set of approximate norms that can be optimally monitored given a set of monitors and queries, therefore approaching the problem from another perspective.

Alechina *et al.* [1] define norms using LTL (linear temporal logic) formulas. They introduce the concept of a guard, which uses lookahead mechanisms to detect future norm violations. The size of the lookahead window is bounded to reduce the amount of computation in the future (they have complete knowledge of the past), and have similarities with the concept of monitor cost in our work. The main difference is that, while we use a combination of monitors to be able to detect a norm violation, they increase their lookahead window size to increase their monitoring capabilities, also increasing its computational cost.

Finally, our work is also similar to that of Bulling *et al.* [3]; both include the concept of monitors and combination of monitors. While we use a relatively simple norm formalism and optimize the cost to monitor these norms, their specification uses LTL-formulas, focusing on its properties and relations. Their current framework does not include norms or monitors costs.

7 Conclusion and Future Work

In this paper we extend the state of the art in norm monitoring by dropping the assumption that a monitor system has full observability, i.e., that monitors can observe all actions performed. Adding the notion of a set of available monitors and an associated cost results in the problem of finding a monitor placement in order to maximize the number of detectable violations. While brute-force solutions are impractical because the possible solution search space is exponential in the size of input, we propose heuristics that use a mapping between norms and monitors to find approximate solutions. Our empirical of runtime performance and accuracy shows that these heuristics are both practical in computational terms and approach optimal performance for many realistic domains from the planning literature.

We aim to extend the work in at least three ways. The first is to allow monitors to dynamically modify their placement during execution. In this setting, monitors would be able to observe agent actions and move to locations where more norm violations occur. The second extension would be to consider different agents being able to perform concurrent actions, and how to build monitors able to correctly identify which agent violated a given norm. The third is to allow more complex expressions representing both what monitors can observe and how monitors can be combined, as currently we only consider conjunctions of monitors. This can increase the richness of our approach, and we intend to investigate heuristics for the use of such more expressive monitors.

Acknowledgments

This work is partially supported by grants from CNPq/Brazil numbers 132339/2016-1 and 305969/2016-1.

References

1. Alechina, N., Bulling, N., Dastani, M., Logan, B.: Practical run-time norm enforcement with bounded lookahead. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems. pp. 443–451 (2015)
2. Alechina, N., Dastani, M., Logan, B.: Norm approximation for imperfect monitors. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multiagent Systems. pp. 117–124 (2014)
3. Bulling, N., Dastani, M., Knobbout, M.: Monitoring norm violations in multi-agent systems. In: Proceedings of the 2013 International Conference on Autonomous Agents and Multiagent Systems. pp. 491–498 (2013)
4. Criado, N.: A practical resource-constrained norm monitor. In: Proceedings of the 2017 International Conference on Autonomous Agents and Multiagent Systems. pp. 1508–1510 (2017)
5. Dignum, F.: Autonomous agents with norms. Artificial Intelligence and Law 7(1), 69–79 (1999)
6. Esteva, M., Rosell, B., Rodriguez-Aguilar, J.A., Arcos, J.L.: Ameli: An agent-based middleware for electronic institutions. In: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems. pp. 236–243 (2004)
7. García-Camino, A., Noriega, P., Rodríguez-Aguilar, J.A.: Implementing norms in electronic institutions. In: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems. pp. 667–673 (2005)
8. García-Camino, A., Rodríguez-Aguilar, J.A., Sierra, C., Vasconcelos, W.: Constraint rule-based programming of norms for electronic institutions. In: Proceedings of the 2009 International Conference on Autonomous Agents and Multiagent Systems. pp. 186–217 (2009)
9. Ghallab, M., Nau, D., Traverso, P.: Automated planning: theory & practice. Elsevier (2004)
10. Luck, M., d'Inverno, M.: Constraining autonomy through norms. In: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems. pp. 674–681 (2002)
11. Modgil, S., Faci, N., Meneguzzi, F., Oren, N., Miles, S., Luck, M.: A framework for monitoring agent-based normative systems. In: Proceedings of the 2009 International Conference on Autonomous Agents and Multiagent Systems. pp. 153–160 (2009)
12. Oren, N., Panagiotidi, S., Vázquez-Salceda, J., Modgil, S., Luck, M., Miles, S.: Towards a formalisation of electronic contracting environments. In: Coordination, Organizations, Institutions and Norms in Agent Systems IV, pp. 156–171 (2009)
13. Ossowski, S.: Agreement technologies, vol. 8. Springer Science & Business Media (2012)
14. Savarimuthu, B.T.R., Cranefield, S.: Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. Multiagent and Grid Systems 7(1), 21–54 (2011)
15. Sutinen, J.G., Andersen, P.: The economics of fisheries law enforcement. Land economics 61(4), 387–397 (1985)

The Perils of Classifying Political Orientation From Text

Hao Yan, Allen Lavoie*, and Sanmay Das

Washington University in St. Louis, St. Louis, USA
{haoyan,allenlavoie,sanmay}@wustl.edu

Abstract. Political communication often takes complex linguistic forms. Understanding political ideology from text is an important methodological task in studying political interactions between people in both new and traditional media. Therefore, there has been a spate of recent research that either relies on, or proposes new methodology for, the classification of political ideology from text data. In this paper, we study the effectiveness of these techniques for classifying ideology in the context of US politics. We construct three different datasets of conservative and liberal English texts from (1) the congressional record, (2) prominent conservative and liberal media websites, and (3) conservative and liberal wikis, and apply text classification algorithms with a domain adaptation technique. Our results are surprisingly negative. We find that the cross-domain learning performance, benchmarking the ability to generalize from one of these datasets to another, is poor, even though the algorithms perform very well in within-dataset cross-validation tests. We provide evidence that the poor performance is due to differences in the concepts that generate the true labels across datasets, rather than to a failure of domain adaptation methods. Our results suggest the need for extreme caution in interpreting the results of machine learning methodologies for classification of political text across domains. The one exception to our strongly negative results is that the classification methods show some ability to generalize from the congressional record to media websites. We show that this is likely because of the temporal movement of the use of specific phrases from politicians to the media.

1 Introduction

Political discourse is a fundamental aspect of government across the world, especially so in democratic institutions. In the US alone, billions of dollars are spent annually on political lobbying and advertising, and language is carefully crafted to influence the public or lawmakers [10, 11]. Matthew Gentzkow won the John Bates Clark Medal in economics in 2014 in part for his contributions to understanding the drivers of media “slant.” With the increasing prevalence of social media, where activity patterns are correlated with political ideologies [2], companies are also striving to identify users’ ideologies based on their comments on political issues, so that they can recommend specific news and advertisements to them.

The manner in which political speech is crafted and words are used creates difficulties applying standard methods. Political ideology classification is a difficult task

* Now at Google Brain

even for people – only those who have substantial experience in politics can correctly classify the ideology behind given articles or sentences. In many political ideology labeling tasks, it is even more essential than in tasks that could be thought of as similar (e.g. labeling images, or identifying positive or negative sentiment in text) to ensure that labelers are qualified before using the labels they generate [5, 21].

One of the reasons why classification of political texts for inexperienced people is hard is because different sides of the political spectrum use slightly different terminology for concepts that are semantically the same. For example, in the US debate over privatizing social security, liberals typically used the phrase "private accounts" whereas conservatives preferred "personal accounts" [13]. Nevertheless, it is well-recognized that "dictionary based" methods for classifying political text have trouble generalizing across different domains of text [17].

Many methods based on machine learning techniques have also been proposed for the problem of classifying political ideology from text [1, 21, 23]. The training and testing process typically follows the standard validation rules: first split the dataset into a training set and a test set, then propose an algorithm and train a classification model based on the training set and finally test on the test set. These methods have been achieving increasingly impressive results, and so it is natural to assume that classifiers trained to recognize political ideology on labeled data from one type of text can be applied to different types of text, as has been common in the social science literature (e.g. Gentzkow and Shapiro using phrases from the Congressional Record to measure the slant of news media [13], or Groseclose and Milyo using citations of different think tanks by politicians to also measure media bias [18]). However, these papers are classifying the bias of entire outlets (for example, *The New York Times* or *The Wall Street Journal*) rather than individual pieces of writing, like articles. Such generalization ability is not obvious in the context of machine learning methods working with smaller portions of text, and must be put to the test.

The main question we ask in this paper is whether the increasingly excellent performance of machine learning models in cross-validation settings will generalize to the task of classifying political ideology in text generated from a *different* source. For example, can a political ideology classifier trained on text from the congressional record successfully distinguish between liberal and conservative news articles? One immediate problem we face in engaging this question is the absence of large datasets with political ideology labels attached to individual pieces of writing. Therefore, we assemble three datasets with very different types of political text and an easy way of attributing labels to texts. The first is the congressional record, where texts can be labeled by the party of the speaker. The second is a dataset of articles from two popular web-based publications, `Townhall.com`, which features conservative columnists, and `salon.com`, which features liberal writers. The third is a dataset of political articles taken from Conservapedia (a conservative response to Wikipedia) and RationalWiki (a liberal response to Conservapedia). In each of these cases there is a natural label associated with each article, and it is relatively uncontroversial that the labels align with common notions of liberal and conservative. We show that standard classification techniques can achieve high performance in distinguishing liberal and conservative pieces of writing in cross-validation experiments on these datasets.

It is tempting to assume that there is enough shared language across datasets that one can generalize from one to the other for new tasks, for example, for detecting bias in Wikipedia editors, or the political orientation of op-ed columnists. However, is it really reasonable to extrapolate from any of these datasets to others? As a motivating example, we show that the results of training bag-of-bigram linear classifiers using the three different datasets above and then using them to identify the political biases of Wikipedia administrators leads to wildly inconsistent results, with virtually no correlation between the partisanship rankings of the administrators based on the three different training sets. More generally, we show that, with one exception, the unaltered cross-domain performance of different classifiers on these datasets is abysmal, and there is only marginal benefit from applying a state-of-the-art domain adaptation technique (marginalized stacked denoising autoencoders [6]). The exception is in using data from the congressional record to predict whether articles are from Salon or Townhall, consistent with Gentzkow and Shapiro's results on media bias. A temporal analysis suggests that this is because phrases move in a rapid and predictable way from the congressional record to the news media. However, even in this domain, we provide evidence that the underlying concepts (Salon vs. Townhall compared with Democrat vs. Republican) are significantly different: adding additional labeled data from one domain actively hurts performance on the other. Our results are robust to using regressions on measures of political ideology (DW-Nominate scores [26]) rather than simple classifications of partisanship. Our overall results suggest that we should proceed with extreme caution in using machine learning (or phrase-counting) approaches for classifying political text, especially in situations where we are generalizing from one type of political speech to another.

1.1 Related Work.

While our methods and results are general, we focus in this paper on political ideology in the US context, since there is already a rich literature on the topic, as well as abundant data. Political ideology in U.S. media has been well studied in economics and other social sciences. Groseclose *et al.,* [18] calculate and compare the number of times that think tanks and policy groups were cited by mainstream media and congresspeople. Gentzkow *et al.,* [13] generate a partisan phrase list based on the Congressional Record and compute an index of partisanship for U.S. newspapers based on the frequency of these partisan phrases. Budak *et al.,* [5] use Amazon Mechanical Turk to manually rate articles from major media outlets. They use machine learning methods (logistic regression and SVMs) to identify whether articles are political news, but then use human workers to identify political ideology in order to determine media bias. Ho *et al.,* [20] examine editorials from major newspapers regarding U.S. Supreme Court cases and apply the statistical model proposed by Clinton *et al.,* [7]. All of the above research gives us quantitative political slant measurements of U.S. mainstream media outlets. However, these political ideology classification results are corpus-level rather than article level or sentence level.

The machine learning community has focused more on the learning techniques themselves. Gerrish *et al.,* [14] propose several learning models to predict voting patterns. They evaluate their model via cross-validation on legislative data. Iyyer *et al.,* [21] apply recursive neural networks in political ideology classification. They use Convote [30] and the Ideological Books Corpus [19]. They present cross-validation results and do not analyze performance on different types of data. Ahmed *et al.,* [1] propose an LDA-based topic model to estimate political ideology. They treat the generation of words as an interaction between topic and ideology. They describe an experiment where they train their model based on four blogs and test on two new blogs. However, political blogs are considerably less diverse than our datasets; since the articles in our datasets are generated in completely different ways (speeches, crowdsourcing and editorials). The results in this paper constitute a more general test of cross-domain political ideology learning.

Cross-domain text classification methods are an active area of research. Glorot *et al.,* [15] propose an algorithm based on stacked denoising autoencoders (SDA) to learn domain-invariant feature representations. Chen *et al.,* [6] come up with a marginalized closed-form solution, mSDA. Recently, Ganin *et al.,* [12] have proposed a promising "Y" structure end-to-end domain adversarial learning network, which can be applied in multiple cross-domain learning tasks.

Cohen *et al.,* [8] investigate the classification of political leaning across three different groups (based on activity level) of Twitter users. Without any domain adaptation methodology, they show that cross-domain classification accuracy declines significantly compared with in-domain accuracy. Our work provides a view across much more diverse data sources than just social media, and engages the question of domain adaptation more substantively.

2 Data and Methods

2.1 Data

Mainstream newspapers and websites have been widely used in political ideology research [3, 5, 13]. However, these datasets contain many non-political articles, and the political articles in news datasets are typically non-partisan [5]. Therefore, we carefully construct three datasets that we expect to be partisan: (1) The Congressional Record, containing statements by members of the Republican and Democratic parties in the US congress; (2) News media articles from Salon (a left-leaning website) & Townhall (a right-leaning one); and (3) Articles related to American politics from two collectively constructed "new media" websites, Conservapedia (conservative) & RationalWiki (liberal). Details of the construction process and the resulting corpora are in the appendix.

2.2 Methods

Text Preprocessing We perform some preprocessing on all the datasets to extract content rather than references and metadata, and also standardize the text by lowercasing, stemming, removing stopwords and other extremely common and venue-specific words.

Logistic Regression Models Logistic regression is a standard and useful technique for text classification. We extract bigrams from the text and Term Frequency-Inverse

Document Frequency weighting to construct the feature representation for logistic regression to use (and denote the overall method TF-IDFLR in what follows). We use the implementation provided in the scikit-learn machine learning package [25].

Marginalized Stacked Denoising Autoencoders for domain adaptation Marginalized Stacked Denoising Autoencoders (mSDA) [6] are a state-of-the-art cross-domain text classification method [12]. Given bag-of-words input of text from two different domains, mSDA provides a closed-form representation of the input, and is faster than the original Stacked Denoising Autoencoder (SDA) [15] without loss of classification accuracy. We use TF-IDF bag-of-bigrams vectors as the input to mSDA, the original mSDA Python package[1] for the implementation of mSDA in combination with the logistic regressions described above in our domain adaptation experiments.

Semi-Supervised Recursive Autoencoders Recently, there have been rapid advances in text sentiment and ideology classification based on recursive neural networks. Most of this work is based on sentence or phrase level classification. Some of these methods use fully labeled [29] or partially labeled [21] parsed sentence trees, and some need large numbers of parameters [27, 29]. Since we have large datasets available to use, we use semi-supervised recursive autoencoders (RAE) [28], which do not need parse trees, labels for all nodes in the parse trees, or a large number of parameters. We use the MATLAB package distributed by Socher *et al.,* [28][2]. We do not transform the words down to their linguistic roots when we apply the RAE method since we need to use a word dictionary.

3 Results

3.1 Cross-domain consistency

The first question is whether training on different domains yields consistent results in classifying political ideology. We evaluate this on a motivating task that is exactly the type of task that one may wish to use these types of tools for, determining ideological bias among Wikipedia administrators.

For each of the 500 most active Wikipedia administrators, we concatenate all the strings they have added to pages on Wikipedia related to U.S. politics and classify the resulting "body of work" of that administrator using the three different training sets (the Congressional Record is #1, Salon/Townhall is #2, and RationalWiki/Conservapedia is #3). Each classifier produces a ranking of these 500 administrators. Shockingly we find that these rankings have **virtually no correlation with each other** (see Table 1).

Somewhat more anecdotally, we can also look at the ranks of some users from each method. We select the three most liberal users according to each of the three classifiers and find their positions in the other two lists. The results are in Table 2 and again demonstrate how diverse the rankings can be based on the training sets.

3.2 Consistency across time

[1] `http://www.cse.wustl.edu/~kilian/code/files/mSDA.zip`

[2] `http://nlp.stanford.edu/~socherr/codeDataMoviesEMNLP.zip`

User Sorted Lists	Spearman's ρ	Kendall's τ
U_1, U_2	-0.004588	-0.003469
U_2, U_3	0.005201	0.002133
U_3, U_1	-0.073204	-0.048652

Table 1: Correlation between the user ideology ranks as determined by the three different training sets. U_1 is the rank vector based on the classifier trained on Congressional Record, U_2 is based on Salon / Townhall and U_3 is based on RationalWiki / Conservapedia. Both ρ and τ are close to 0, demonstrating almost no correlation (the statistics range from -1 for perfectly anticorrelated to +1 for perfectly correlated).

User Name	U_1	U_2	U_3
Barek	1	282	487
ERcheck	2	387	35
Widr	3	345	496
James086	262	1	356
Penwhale	455	2	300
Dave souza	97	3	240
Gyrofrog	425	141	1
Smartse	416	358	2
Rigadoun	418	38	3

Table 2: Rankings of the three most liberal users according to classifiers trained on each of the training sets.

The words used to describe politics change across time, as do the topics of importance. Therefore, political articles that are distant in time from each other will be less similar than those written during the same period. We now study whether this is a significant issue for the logistic regression methods by focusing on the Salon and Townhall dataset. We use 2006 Salon and Townhall articles as a training set and future years (from 2007 to 2014) as separate test sets.

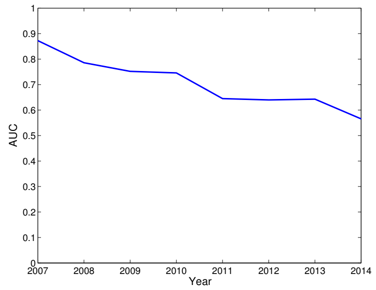


Fig. 1: Salon & Townhall year-based timeline test. The training set is 2006 Salon & Townhall data. The test sets are individual year data from 2007 to 2014, also from Salon & Townhall.

Figure 1 shows the AUC across time. The AUC for 2007 is 0.872, which means that the Salon & Townhall articles in 2006 and 2007 are similar enough for successful generalization of the ideology classifier from one to the other. However, the prediction accuracy goes down significantly as the dates of the test set become further out in the future, as the nature of the discourse changes. It is now clear that our classification methods have generalization problems both across domains and across time.

3.3 Domain adaptation

Now we turn to a more comprehensive analysis. We examine the performance of several different methods across the three labeled datasets. We study linear classifiers and recursive autoencoders as described above, as well as the mSDA method for domain adaptation. In order to account for the effects of time-varying language use demonstrated above, we restrict our methods to train and test only on data from the same year, and then aggregate results across years.

Training Set \ Test Set	Congressional Record	Salon & Townhall	Conservapedia & RationalWiki
Congressional Record	0.8299 (TF-IDFLR) 0.8136 (RAE)	**0.6935**(mSDA) 0.6731 (TF-IDFLR) 0.5937(RAE)	0.4729(mSDA) **0.4940** (TF-IDFLR) 0.4655 (RAE)

Salon & Townhall	**0.6038**(mSDA)
	0.5861 (TF-IDFLR)
	0.5363 (RAE)
		0.9193(TF-IDFLR)
		0.9041(RAE)
			0.5234(mSDA)
			0.5080 (TF-IDFLR)
			0.5527(RAE)

Conservapedia & RationalWiki	**0.5260**(mSDA) 0.5012 (TF-IDFLR) 0.4674 (RAE)	**0.5835**(mSDA) 0.5282 (TF-IDFLR) 0.5711 (RAE)	0.8493 (TF-IDFLR) 0.8180 (RAE)

Table 3: Domain adaptation test based on three data sets

Table 3 shows the average AUC for each group of experiments. The within-domain cross-validation results (on the diagonal) are excellent for both the linear classifier and the RAE. However, the naive cross-domain generalization results are uniformly terrible, often barely above chance. While we could hope that using a sophisticated domain-adaptation technique like mSDA would help, the results are disappointing: in only one cross-domain task (generalizing from the Congressional Record to Salon and Townhall) does it help to achieve a reasonable level of accuracy. The AUC score gaps between cross-validation and domain adaptation results indicate that, even with a state-of-the-art domain adaptation algorithm, cross-text domain political ideology identification is not, at this point, able to give reliable results. It is of note that the best performance is in generalizing from the congressional record to a media dataset (Salon/Townhall) because it adds weight to the existing line of research starting from Gentzkow and Shapiro on how language flows from politicians to the media. (Implementation details and parameter choices for Sections 3.1-3.3 can be found in the appendix)

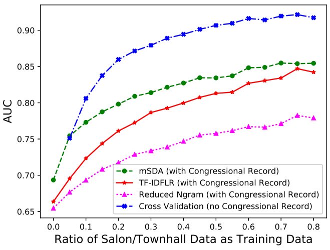


Fig. 2: AUC on Salon/Townhall as a function of the proportion of the labeled (Salon/Townhall) dataset used in training. The results show that including labeled data from the Congressional Record never helps and actively hurts classification accuracy in almost all settings, and that restricting features to ngrams with sufficient support in both datasets does not help either.

3.4 Failure of domain adaptation, or distinct concepts?

There are two plausible hypotheses that could explain these negative results. H1: The domain adaptation algorithm algorithm is failing (probably because it is easy to overfit labeled data from any of the specific domains), or H2: The specific concepts we are trying to learn are actually different or inconsistent across the different datasets. We perform several experiments to try and provide evidence to distinguish between these hypotheses. First, we may be able to reduce overfitting by restricting the features to ngrams that have sufficient support (operationally, at least 5 appearances) in both sets of data (this reduces the dimensionality of the space and would lead to a greater likelihood of the “true” liberal/conservative concept being found if there were many accurate hypotheses that could work in any individual dataset). Second, we can examine performance as we include more and more *labeled* data from the target domain in the training set. In the limit, if the concepts are consistent, we would not expect to see any degradation in (cross-validation) performance on the source domain from including labeled data from the target domain in training.

We focus on the Salon/Townhall and Congressional Record data sets here since they are the most promising for the possibility of domain adaptation. We combine part of the Salon/Townhall data with Congressional Record as training set. Then we use the rest of the Salon/Townhall data set as the test set, increasing the percentage of the Salon/Townhall dataset used in training from 0% to 80%, and compare with cross-validation performance on just the Salon/Townhall dataset.

Figure 2 shows that including labeled data from the Congressional Record never helps and, once we have at least 10% of labels, actively hurts classification accuracy on the Salon/Townhall dataset. Restricting to bigrams that appear in both datasets at least 5 times further degrades the performance. This demonstrates quite clearly that the problem is not overfitting a specific dataset when there are many correct concepts available, it is that the concept of being from Salon or Townhall is significantly different than the concept of being from a Democratic or Republican speech. Therefore, the hope of successful domain-agnostic classification of political orientation based on text data is significantly diminished.

3.5 Temporal movement of topics

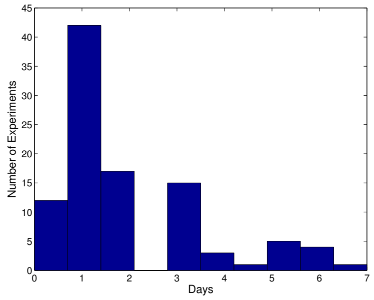


Fig. 3: Distribution of median value of time lag results in each experiment

The silver lining so far is that there is at least some ability to predict the political orientation of web-based news media based on the congressional record. We can further investigate this insight and demonstrate the utility of the data we have collected by examining the question temporally. Leskovec *et al.,* [22] investigated the time lag regarding news events between the mainstream media and blogs. We ask a similar question – who discusses “new” political topics in the first place – congress or the media?

In order to answer this question, we examine mutual trigrams in the Congressional Record and Salon&Townhall datasets. We find all new trigrams in any given year (those which did not appear in the previous year and appeared at least twice in the media data and five times in the congressional record in the given year and the next one), and then construct the time lags between first appearance in each of the two datasets, excluding congressional recess days. Since the congressional record is much larger, we subsample and repeat the experiment many times to get a distribution of time lags.

In each of these bootstrapped samples, there is a median time lag between the first appearance of a phrase in the congressional record and its first appearance in the media dataset. Figure 3 shows the distribution of these medians. The median is never negative, and is on average 2 days, showing a definite tendency for phrases to travel from the congressional record to the media rather than the other way round. The entire distribution also shows a slight bias towards the media picking up on congressional topics of discussion after the fact. These results help to explain the relative success of domain adaptation from the congressional record to the media dataset.

4 Conclusion

Text analytics is becoming a central methodological tool in analyzing political communication in many different contexts. It is obviously very valuable to have a good way of measuring political ideology based on text. Our work sounds a cautionary note in this regard by demonstrating the difficulty of classifying political text across different contexts. We provide strong evidence that, in spite of the fact that writers or speech makers in different domains often self-identify or can be relatively easily identified by humans as being conservative or liberal, the concepts are distinct enough across datasets (even in just the US political context!) that generalization is extremely difficult. We note that, while we have presented our results in the context of classification, we get identical results when using measures of political ideology on a real-valued spectrum (the standard DW-Nominate score [26]) as the target of a regression task (this is only feasible for the congressional record, since the scores of congresspeople can be obtained as a function of their voting record). Our results demonstrate the need for extreme caution in the application of machine learning techniques to classifying political ideologies, especially when such efforts are made across domains.

References

1. Ahmed, A., Xing, E.P.: Staying informed: Supervised and semi-supervised multi-view topical analysis of ideological perspective. In: Proc. EMNLP. pp. 1140–1150 (2010)
2. Bakshy, E., Messing, S., Adamic, L.A.: Exposure to ideologically diverse news and opinion on Facebook. Science 348(6239), 1130–1132 (2015)
3. Baum, M.A., Groeling, T.: New media and the polarization of American political discourse. Polit. Comm. 25(4), 345–365 (2008)
4. Brown, A.R.: Wikipedia as a data source for political scientists: Accuracy and completeness of coverage. PS: Polit. Sci. & Politics 44(02), 339–343 (2011)
5. Budak, C., Goel, S., Rao, J.M.: Fair and balanced? Quantifying media bias through crowdsourced content analysis. Public Opin. Quarterly 80(S1), 250–271 (2016)
6. Chen, M., Weinberger, K.Q., Xu, Z., Sha, F.: Marginalized stacked denoising autoencoders for domain adaptation. In: Proc. ICML. pp. 767—-774 (2012)
7. Clinton, J., Jackman, S., Rivers, D.: The statistical analysis of roll call data. Am. Polit. Sci. Rev. 98(02), 355–370 (2004)

8. Cohen, R., Ruths, D.: Classifying political orientation on twitter: It's not easy! In: Proc. ICWSM (2013)
9. Das, S., Lavoie, A., Magdon-Ismail, M.: Manipulation among the arbiters of collective intelligence: How Wikipedia administrators mold public opinion. ACM Trans. on the Web 10(4), 24:1–24:25 (2016)
10. DellaVigna, S., Kaplan, E.: The Fox News effect: Media bias and voting. Q. J. Econ. 122(3), 1187–1234 (2007)
11. Entman, R.M.: How the media affect what people think: An information processing approach. J. Politics 51(02), 347–370 (1989)
12. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. JMLR 17(59), 1–35 (2016)
13. Gentzkow, M., Shapiro, J.M.: What drives media slant? Evidence from U.S. daily newspapers. Econometrica 78(1), 35–71 (2010)
14. Gerrish, S., Blei, D.M.: Predicting legislative roll calls from text. In: Proc. ICML. pp. 489–496 (2011)
15. Glorot, X., Bordes, A., Bengio, Y.: Domain adaptation for large-scale sentiment classification: A deep learning approach. In: Proc. ICML. pp. 513–520 (2011)
16. Greenstein, S., Zhu, F.: Is Wikipedia biased? Am. Econ. Rev. 102(3), 343–348 (2012)
17. Grimmer, J., Stewart, B.M.: Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political Analysis pp. 1–31 (2013)
18. Groseclose, T., Milyo, J.: A measure of media bias. Q. J. Econ. 120(4), 1191–1237 (2005)
19. Gross, J., Acree, B., Sim, Y., Smith, N.A.: Testing the Etch-a-Sketch hypothesis: A computational analysis of Mitt Romney's ideological makeover during the 2012 primary vs. general elections. In: APSA Annual Meeting (2013)
20. Ho, D.E., Quinn, K.M.: Measuring explicit political positions of media. Q. J. Polit. Sci. 3(4), 353–377 (2008)
21. Iyyer, M., Enns, P., Boyd-Graber, J., Resnik, P.: Political ideology detection using recursive neural networks. In: ACL. pp. 1113–1122 (2014)
22. Leskovec, J., Backstrom, L., Kleinberg, J.: Meme-tracking and the dynamics of the news cycle. In: Proc. KDD. pp. 497–506. ACM (2009)
23. Lin, W., Xing, E.P., Hauptmann, A.G.: A joint topic and perspective model for ideological discourse. In: Proc. ECML-PKDD. pp. 17–32 (2008)
24. Mikolov, T., Chen, K., Corrado, G.S., Dean, J.: Efficient estimation of word representations in vector space. In: Proc. ICLR (2013), http://arxiv.org/abs/1301.3781
25. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. JMLR 12, 2825–2830 (October 2011)
26. Poole, K.T., Rosenthal, H.: Congress: A political-economic history of roll call voting. Oxford University Press (1997)
27. Socher, R., Huval, B., Manning, C.D., Ng, A.Y.: Semantic compositionality through recursive matrix-vector spaces. In: Proc. EMNLP. pp. 1201–1211 (2012)
28. Socher, R., Pennington, J., Huang, E.H., Ng, A.Y., Manning, C.D.: Semi-supervised recursive autoencoders for predicting sentiment distributions. In: Proc. EMNLP. pp. 151–161 (2011)
29. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proc. EMNLP. pp. 1631–1642 (2013)
30. Thomas, M., Pang, B., Lee, L.: Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In: Proc. EMNLP. pp. 327–335 (2006)

Appendix

A Datasets

A.1 Congressional Record.

The U.S. Congressional Record preserves the activities of the House and Senate, including every debate, bill, and announcement. We use the party affiliation of the speaker (Democrat or Republican) as an indication of ideology (liberal or conservative). We retrieve the floor proceedings of both the Senate and House from 2005 to 2014. We separate the proceedings into segments with a single speaker. For each of these segments, we extract the speaker and their party affiliation (Democrat, Republican or independent)In order to focus on partisan language, we excluded speech from independents, and from clerks and presiding officers.

A.2 Salon and Townhall.

We collect articles tagged with “politics” from Salon, a website with a progressive/liberal ideology, and all articles from Townhall, which mainly publishes reports about U.S. political events and political commentary from a conservative viewpoint.

A.3 Conservapedia and RationalWiki.

Conservapedia (`http://www.conservapedia.com/`) is a wiki encyclopedia project website. Conservapedia strives for a conservative point of view, created as a reaction to what was seen as a liberal point of view from Wikipedia. RationalWiki (`http://rationalwiki.org/`) is also a wiki encyclopedia project website, which was, in turn, created as a liberal response to Conservapedia. RationalWiki and Conservapedia are based on the MediaWiki system. Once a page is set up, other users can revise it. For RationalWiki, we download pages ranking in the top 10000 in number of revisions. We further select pages whose categories contain the following word stems: *liber, conserv, govern, tea party, politic, left-wing, right-wing, president, u.s. cabinet, united states senat, united states house*. Because the Conservapedia community has more articles than RationalWiki, we download the top 40000 pages. We apply the same political keywords list we use for RationalWiki. We always use the last revision of any page for a given time period.

Table 4 shows the counts of articles in the liberal and conservative parts of each of the three datasets by year. Our datasets have the following properties that make them useful for political ideology learning and evaluation in the context of U.S. politics:

- The content is selected to be relevant to U.S. politics.
- The content can predictably be labeled as conservative or liberal by a somewhat knowledgeable human. While it is true that not all speeches by Democrats are liberal, and not all articles on Townhall conservative, since these are subjectively defined, this is nevertheless as clean a delineation as we can hope for.
- The creation times of items in the three datasets have substantial overlap;

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Democrat (CR)	14504	11134	17990	11053	14580	11080	11161	8540	9673	7956	0	0
Republican (CR)	11478	9289	12897	8362	13351	7878	9141	6841	8212	6585	0	0
Salon	1613	1561	2161	2598	2615	1650	1860	1630	865	123	0	0
Townhall	27	143	290	341	174	176	258	380	441	674	0	0
RationalWiki	0	0	302	514	666	854	1086	1208	1342	1402	1480	1480
Conservapedia	0	93	1752	2381	2933	3214	3467	3698	3792	3863	3937	3938

Table 4: Article distributions by year in the three datasets. Democrat (CR), Salon, and RationalWiki are assumed to be liberal, while Republican (CR), Townhall, and Conservapedia are assumed to be conservative.

A.4 Wikipedia

We also motivate our task by attempting to classify bias on Wikipedia, an important task [9]. Wikipedia is the largest encyclopedia project in the world and is widely used in both natural language processing and political science studies [4, 24]. Wikipedia is considered to have become nonpartisan as many users have contributed to political entries [16]. We focus on edits made by admins on political topics in Wikipedia. We download the English Wikipedia dump from March 4, 2015. To focus on US politics, we extract all articles (with full edit history) that belong to WikiProject United States[3] and satisfy the same political keywords requirement that we use for RationalWiki, yielding 4659 articles in total. We then collect all edits added or subtracted by each active Wikipedia admin.

B Details of Experimental Methodology

B.1 Cross-domain consistency

For the Congressional Record and Salon/Townhall datasets, we use data from 2005 to 2014. For the RationalWiki/Conservapedia datasets, we use the data from 2014 as capturing a recent snapshot. For this dataset only we use feature hashing to project the bigram features into a lower dimensional non-sparse feature space. We set the dimension of the hashed vector $n\_features = 20000$, $ngram\_range = (2, 2)$, and $decode\_error = ignore$. We use a so-called "balanced" logistic regression classifier to deal with the problem of class imbalance. All other parameters are the defaults in the scikit-learn package for both feature hashing vectorizer and logistic regression classifier.

B.2 Consistency across time

We use the TF-IDFLR method for this experiment. For the vectorizer, we set $min\_df = 5$, $ngram\_range = (2, 2)$ and $decode\_error = ignore$. For logistic regression classifier, we set $class\_weight = balanced$ to re-weight training samples. Other parameters are set to the default values in the scikit-learn package.

[3] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_United_States

B.3 Domain adaptation

The linear classifier is the TF-IDFLR method described above. The RAE algorithm trains embeddings using sentences subsampled from the data in order to balance conservative and liberal training sentences, and then a logistic regression classifier is used on top of the embeddings thus trained. The marginalized stacked denoising autoencoder, which is expected to find features that convey domain-invariant political ideology information, is run on TF-IDF bigram features before a logistic regression is applied on top of that feature representation. We use five-fold cross validation when the training and testing sets are the same.

Democracy Models and Civic Technologies: Tensions, Trilemmas, and Trade-offs

Marta Poblet[1] and Enric Plaza[2]

[1] RMIT University, Melbourne, VIC 3000, Australia
[2] Artificial Intelligence Institute, Spanish Research Council, Bellaterra 08193 Spain
marta.pobletbalcell@rmit.edu.au
enric@iiia.csic.es

Abstract. This paper aims at connecting democratic theory with civic technologies in order to highlight the links between some theoretical tensions and trilemmas and design trade-offs. First, it reviews some tensions and 'trilemmas' raised by political philosophers and democratic theorists. Second, it considers both the role and the limitations of civic technologies in mitigating these tensions and trilemmas. Third, it proposes to adopt a meso-level approach, in between the macro-level of democratic theories and the micro-level of tools, to situate the interplay between people, digital technologies, and data.

Keywords: Democracy, civic technologies, representation, participation, deliberation, linked democracy.

1 Introduction

Over the last two decades, digital technologies have opened up new paths for civic engagement and political participation. Hundreds of websites, portals, platforms and mobile apps enable citizens across the globe to organise campaigns, vote initiatives and sign petitions, monitor their representatives and track parliamentary activity, propose ideas and draft legislation or constitutions. Governments at different levels adopt digital technologies to develop 'open government' and 'open data' strategies to promote citizens' participation and increase transparency. Crowdsourcing is now a pervasive method to collect data, information, ideas, and legislative proposals. A growing literature based on case studies and empirical testing provides the basis for further refinement of methods: e.g. smart crowdsourcing [17], expert crowdsourcing [5,6] microtasking [10].

The exploration of new technologies and methods to harness the potential of crowdsourcing for civic action and politics, nevertheless, contrasts with the scarce attention given to the underlying assumptions about democracy, participation, equality, representation, and citizenship. Surprisingly enough, there has been little dialogue between theorists of democracy and citizenship, on the one hand, and digital technologists, information systems and AI experts, on the other, on how civic technologies may redefine our current notions of democracy, participation, equality, representation, and citizenship.

Our goal in this paper is threefold. First, we aim to induce a discussion on how to reinterpret some of these notions by reviewing some tensions and 'trilemmas' raised by political philosophers and democratic theorists. Second, we consider both the role and the limitations of civic technologies in mitigating these tensions and trilemmas. Third, we propose to adopt a meso-level approach, in between the macro-level of democratic theories and the micro-level of tools, to situate the interplay between people, digital technologies, and data. As different groups in different social contexts use digital tools and data differently, it is at this meso level that we can elucidate the trade-offs with the notions of the trilemmas. We conceptualise the meso-level as the institutional level, for the notion of institution will give us a framework to analyse the use of technology in a given social context.

2 Some Tensions and 'Trilemmas' in Democratic Theory

2.1 A Condorcetian reading of representation

The tensions between key concepts in democratic theory, notably sovereignty, representation, participation, equality, and citizenship have long been debated. In her work on representative democracy, Nadia Urbinati has noted that both Montesquieu and Rousseau were 'the first theorists to argue (for divergent reasons) that an unsolvable tension exists between democracy, sovereignty, and representation' [23, p. 54]. More specifically:

> Montesquieu separated representation from democracy, and Rousseau representation from sovereignty. Montesquieu argued that a state where the people delegated their 'right of sovereignty' could not be democratic and must be classified as a species of mixed government and in fact an aristocracy. Rousseau saw such a state as non-political from the start and illegitimate because the people lost their political liberty along with the power to vote on legislation directly: unless all citizens were lawmakers, there were no citizens at all. In both cases, democracy and sovereignty excluded representation (p. 54).

Urbinati argues that this exclusion remains implicit within contemporary theories of representative government for which "from a theoretical point of view, a 'represented democracy', although technically feasible, is an oxymoron, while direct democracy, although the norm, is impractical" [22, p.55]. Yet, Urbinati denies this incompatibility to be the only legacy of 18th century's political philosophy when it comes to the idea of representation.[1] In supporting her claim for a 'democratic under-

[1] 'Rather than a monolithic entity, the theory of representative government formed, since its birth, a complex and pluralistic family whose democratic wing was not the exclusive property of those who advocated for participation against representation.' [22, p. 55].

standing of representation' she draws on Condorcet's *Plan de Constitution* submitted to the French National Assembly in 1793. Condorcet's proposal, eventually rejected by both his fellow Girondins and the Jacobins, contains what Urbinati describes as 'an institutional order that is one of the most democratically advanced and imaginative Europe has produced in the last two centuries' [22, p. 56]:

> Condorcet's constitution designed a political order that was horizontal and acephalous (parliamentary, not presidential) and rigorously based on the centrality of the legislative power, a power held by a multiplicity of actors and performed in multiple times and within a plurality of spaces. The function of legislation was performed within assemblies – elected assembly and assemblies of the citizens (*assemblées primaires*) – and was held by the representatives along with (not instead of) the citizens who 'enjoyed' both the electoral right and the right to revoke or censure the laws (constitutional and ordinary). [22, p. 59-60].

Condorcet, Urbinati notes, reconciles sovereignty, representation, and participation by making 'citizens' participation essential to both the functioning of representative government and the preservation of political liberty' [22, p. 60]. With a comment that echoes Josiah Ober's vision of the role of citizens in ancient Athens [14, 15], Urbinati sees citizen participation in Condorcet's institutional order as a 'source of stability and of innovation', while representation becomes the political device collecting and filtering knowledge for the public interest [22, p. 60].

In our contemporary democracies, representation has become an even more intricate subject, even at the local level [16]. Urbinati and Warren argue that the complexity of issues and the multiple, overlapping constituencies involved call for the extension of the meaning of representation to include non-electoral forms 'that are capable of representing latent interests, transnational issues, broad values, and discursive positions' [23, p. 407]. Moreover, the Internet has also enabled the emergence of online communities of interest beyond geographical boundaries that have no mechanisms of representation in our political systems [9].

It is our contention that digital technologies and AI can facilitate the channelling of these multifaceted forms of representation in unique ways. But a second 'trilemma' needs to be addressed before considering these options.

2.2 The 'trilemma' of democratic reform

James Fishkin, a leading theorist of deliberative democracy, addresses in one of his papers the key question of how to incorporate public deliberation into constitutional processes [4]. In raising this question he introduces what he refers to as the 'trilemma of democratic reform'. To Fishkin, there are three basic principles internal to the design of democratic institutions: political equality (people's views are counted equally), mass participation (we are all given the opportunity to provide informed consent), and deliberation (we are all given the opportunity to provide opinions and weigh competing arguments).

Fishkin suggests that, under normal conditions, any serious effort to attain any of the two principles inevitably hinders the third, so that we cannot satisfy the three principles simultaneously. For example, if we pursue a process driven by political equality and mass participation we are unlikely to get deliberation into the picture because the incentives for people to become seriously informed and engaged are very low ('audience democracy'). Likewise, we can satisfy the principles of political equality and deliberation if we choose (by lot or by random sampling) a microcosm of deliberators (e.g. Fishkin's Deliberative Polls). This microcosm may be representative of the broader population from which it has been extracted, but then this population will have no voice in the process and therefore the principle of mass participation will not be fulfilled. Finally, we can have a process with mass participation (to some extent) and deliberation. This is what most of the current online crowd-civic platforms provide, but what we gather in this case is a 'self-selected microcosm of deliberators', highly engaged and yet, far from being representative of the broader population (so we would be violating the principle of political equality). Tanja Aitamurto *et al*. [1] have also highlighted the tension between the norm of equal representation in democracy and the self-selection bias of crowdsourcing, suggesting that 'crowdsourcing shouldn't strive for statistical representativeness of the public, otherwise the virtues of crowdsourcing would be compromised and its benefits in crowd work would not be achieved.' [1, p. 1]. Statistical representativeness as a requirement may be a debatable issue, but what is at stake here is the *legitimacy* of crowdsourcing in political practice. We also find a self-selection bias in offline political activity, e.g. in parliamentary elections, where the turnout is usually significantly below 100 per cent of the demos. How self-selection affects legitimacy in a political process is a general issue that political theory needs to address in broader terms. Specifically, if we conceptualize political equality in the classical sense [*isegoria* (equal voice) + *isonomia* (equality of political rights)] self-selection does not necessarily diminish the principle of equality (non-participation is an individual decision).

What should we do if the simultaneous achievement of the three principles is not attainable? Fishkin suggests adopting a pragmatic approach to solve his trilemma. Rather than trying to approximate the ideal, he proposes the design of a second best approach or a proxy (and hence his research program on Deliberative Polling, aiming at both the internal and external validity of the process). Nevertheless, Fishkin acknowledges that this solution may incur a democratic deficit, since the resulting views may not be the actual views of the public [4, p. 253]. To tackle this issue, he proposes a process with sequential strategies (for example, a convention followed by a deliberative microcosm and, finally, a referendum) that, combined, cover the three principles at different stages.

The remaining issue, nevertheless, is that deliberation does not travel well across those stages. Fishkin illustrates what he terms 'the weak link of deliberation' with the example of the Australian 1999 referendum, where two different deliberative bodies (a convention and a deliberative poll) had previously reached the opposite conclusion (pro-republic) with regard to the proposal of an Australian republic [4: p.253-254]. The elaboration of Iceland's Constitution is another recent example of the weak connection between deliberative bodies (in this case, the Constitutional Convention and the Parliament). Fishkin proposes to strengthen this link by organizing a Deliberation Day, where the entire population is convened for one day to engage in deliberation followed by a referendum. To motivate participants, Fishkin estimated that an incentive of $300 per participant would act as an adequate incentive [4, p. 258]. No matter how well designed, though, the costs of such events could be extremely prohibitive for many countries, especially considering how short-lived they would be. The question that remains open is whether there is a role for technology in mitigating the trilemma.

3 Mitigating Democratic Trilemmas

Political philosophy addresses both the tensions and trilemmas in democratic theory and practice with a sophisticated conceptual apparatus. Yet, research on the implications of civic technologies for democracy and democratisation processes is still largely overlooked in both deliberative and epistemic accounts of democracy. This compartmentalisation of knowledge is disadvantageous from both a theoretical and empirical perspective. For example, enabling effective non-electoral forms of representation would require a survey of technology options and 'knowledge of what works and when' [17]. Likewise, a better understanding of the underlying principles, models, and concepts of democratic theory would help to inform the design of civic tools and modulate the frequently inflated expectations placed on them.

Digital platforms facilitate the depth and breadth of participation, lowering the barriers to different forms of participation (without precluding offline participation) and improving the 'open access pattern' of a given social order [13]. They also open up the door to new, meaningful forms of mass deliberation and epistemic outcomes [e.g. 10, 19]. To illustrate this point, in Figures 1 and 2 below we compare two models of democracy:

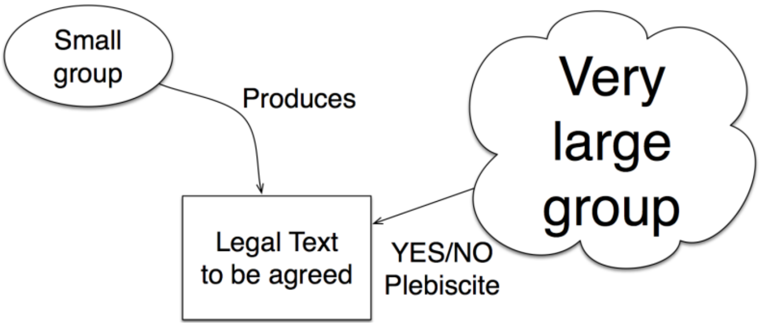


Fig. 1. Plebiscitarian model with deliberative body

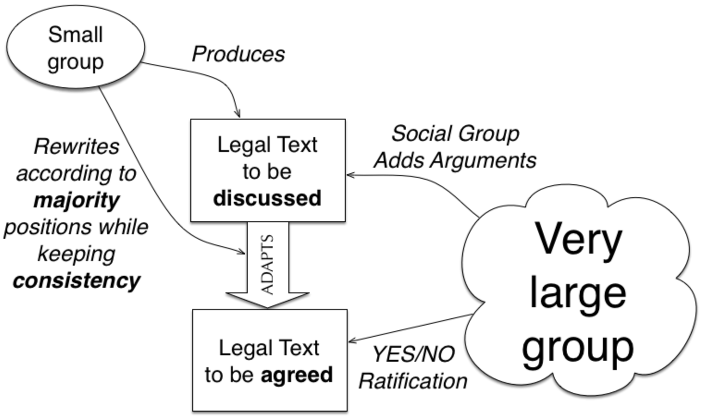

Small group
Produces
Legal Text to be **discussed**
Social Group Adds Arguments
Rewrites according to **majority** positions while keeping **consistency**
ADAPTS
Very large group
Legal Text to be **agreed**
YES/NO Ratification

Fig. 2. Participatory model with deliberative body

Fig. 1 represents a well-known plebiscitarian model of democracy: a small group (for example, a constitutional convention, a parliamentary commission, etc.) produces a legal text. When the text is ready, a referendum is called and citizens can cast a yes/no vote. This model accounts for the principles of political equality, mass participation, and representation. Yet, deliberation is restricted to the small group, as citizens are left with just an *ex post*, binary option (yes/no). Many constitution making processes in Western democracies have followed this path to date.

Fig. 2 visualises a more complex participatory model to mitigate the trilemma. As in the previous case, a small group of people (either drafted by sortition or appointed by some other entity) is given the task of producing a legal text, but in sequential steps. The group deliberates on a first draft, which is open to the general public for comments and suggestions (typically from a self-selected subset of the electorate). The feedback from this very large group is incorporated in the draft and subsequently adapted to produce, after a number of iterations, the text to be agreed and ratified by the electorate. This participatory model was famously deployed in Iceland in 2011, when the meetings and debates of a Constitutional Council of 25 individuals (drafted from a larger pool of citizens) were made publicly available in the Council website for comment via social media and e-mail. The proposal was approved by a two-thirds majority of the voting population in a referendum in late 2012 but it eventually stalled in parliament [8].

Similarly, this model was adopted in Mexico City. On January 2016, the Mayor of the city obtained approval from the federal parliament to initiate a constitution-making process by appointing a group of 30 experts to discuss and draft a proposal.[2] In order to open up the drafting process to the citizenry, the City Council made publicly available a collaborative editing tool for citizens to provide feedback on the specific topics posted by the drafting group. Moreover, as crowdsourced legal drafting does not typically attract a large number of citizens, this approach was complemented

[2] https://www.constitucion.cdmx.gob.mx/constitucion-cdmx/#grupo-trabajo

with other participatory strategies, namely a survey and a Change.org campaign to collect petitions relevant to the constitutional text (at the closing date of the process, 280,678 people had supported 129 petitions). The Constitution of Mexico City was finally published on 5 February 2017, although at the time of writing the Supreme Court of Mexico is hearing a number of appeals to the constitutional text from the federal government, two political parties, and other organisations.[3] Strikingly, both the Icelandic and Mexican constitutional drafts came to a standstill as other institutional bodies were involved. We will review this in Section 3.2 below.

3.1 The technology caveat

Digital platforms have come a long way when it comes to facilitating legal drafting, crowdsourcing of ideas, or structuring large-scale deliberation, but the tasks of aggregating legal and political knowledge for deliberation and decision making remain onerous. In recent years, a number of advances in AI areas such as text mining, argument detection, extraction, and mapping can be applied to support the activity of very large groups, both to improve self awareness (of what they are co-producing) and facilitate knowledge aggregation. Likewise, both small and large groups can benefit from text mining, semantic languages (e.g. RDF, XML), ontologies, linked data, and machine learning when searching, analysing and reusing legal texts to elaborate new ones. For example, using ConstituteProject,[4] constitution makers can now browse nearly 200 constitutions across the world (tagged with more than 300 topical labels) when drafting their own. Global laws are also accessible to law proponents or drafters with services offered, among others, by the World Legal Information Institute[5] or Global Regulation.[6]

To date, online platforms have focused on improving and facilitating mass participation (or at least to include larger numbers of citizens in a political process). Those efforts have proved useful when supporting the participation of dozens, hundreds or, in some cases, thousands of people contributing to an initiative with arguments or comments. Yet, the issue of effectively enabling large-scale, massive participation (that is, hundreds of thousands or even millions of people) is still unresolved.

It is also important to note here the implicit assumption that correlates higher participation with higher legitimacy. Mexico City, to use our previous example, has almost 9 million inhabitants, but what is the threshold for establishing that a constitution crowdsourced from a negligible percentage of its inhabitants is more legitimate than appointing a group of 30 experts? Can future civic technologies really scale up to *mass* participation in elaborating policies and laws, or can legitimacy only be claimed when the crowds are requested to ratify them? Would it be better to design systems that cater for smaller, decentralised, and distributed (offline and online) citizen as-

[3] http://www.eluniversal.com.mx/articulo/nacion/politica/2017/03/10/corte-admite-impugnaciones-contra-constitucion-cdmx

[4] http://constituteproject.org

[5] http://wordlii.org

[6] http://www.global-regulation.com

semblies (thus supporting a renewed version of democratic representation)?[7] While these questions remain open, the answers also depend on political and institutional choices.

3.2 The institutional caveat

A second caveat when trying to mitigate democratic trilemmas is that deploying civic tools for large-scale participation will not guarantee any real influence on either rule making or policy making. As the examples in Iceland and Mexico show, there is no way to ensure that embedding participatory components into the process—regardless of whether this participation is deliberative or not—will eventually have an impact on decision making and, ultimately, will lead to more bottom-up, inclusive decisions.

Over the last two decades, deliberative democrats have set the conditions, procedures, and standards of deliberative processes. More recently, some of them have adopted a 'systemic' approach where some institutions will achieve some principles while others will achieve others, making the institutional system 'deliberative' as a whole [11]. The focus on procedures and standards has also expanded to include the discussion on whether mini-publics (citizen juries, citizen assemblies, deliberative polls, etc.) and other institutional innovations should have a binding force—aligning the outcomes of deliberation with rule or policy making—or have a mere advisory role [e.g. 7]. The debate highlights the underlying tensions between participation and deliberation, but it does so from an abstract perspective. Ironically enough, the discussion on the optimal institutional design to coordinate and translate deliberative outputs at the micro level into aligned policy making is not institutionally anchored. Yet, without such anchoring, it is hard to predict in which particular institutional contexts the new designs will either thrive or languish, and which trade-offs will be required. Empirical studies focusing on the institutional level, such as the Utrecht experiment below, may help to shed some light:

> The key feature of this process of political innovation is that citizens were randomly selected to participate, they received remuneration for their participation and they could be regarded as an alternative form of citizen representation. In contrast with many other forms of participation such as citizen panels, the advice was not 'free': local government had committed beforehand to follow this advice and to translate it to an energy policy plan. Our empirical analysis of this case shows that an interplay between idealist and realist logics explains why they are 'accepted' by the institutionalized democratic system." [12, p. 21].

[7] We also find examples of this option in Buenos Aires, British Columbia, or Ireland. The Swiss 'semi-direct democracy' model [3] is paradigmatic when combining representation and popular sovereignty at the three levels of governance (federal, cantonal and municipal). Approximately four times a year, voting occurs over various issues: federal popular initiatives (constitutional reforms), policies, and election of representatives. Federal, cantonal and municipal issues are polled simultaneously, and the majority of votes are cast by mail.

An intermediate, meso-level approach to both online and offline innovations would help to elucidate the interactions between people, technology, and data in particular settings. It would also provide a framework of analysis to better understand both the emerging properties (and tensions) of these interactions. We have suggested a model of 'linked democracy' to synthetise this framework [2, 18]. Linked democracy, therefore, is a model dynamically linking the distributed interactions between people, data, institutions within both organizational and local contexts.

4 A Proposal for a Meso-level Approach: Some Features

Our proposal consists of analysing political ecosystems where clusters of institutions are distributed throughout with different roles and specialisations, but all connected together in a distributed way. Both the Mexican and Icelandic cases can be analysed through these lenses, as well as, for example, the connected interactions between people, technology and data in a public health ecosystem [2].

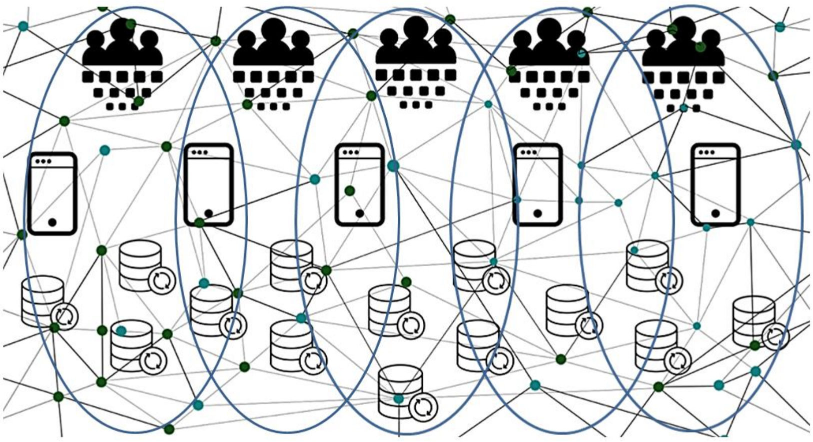


Fig. 3. An ecosystem of linked institutions[8]

It is out of the scope of this paper to present a case study embedded in this meso-level approach. Nevertheless, our proposal here includes outlining the features that may guide such an analysis from the perspective of a linked democracy model. Thus, our analysis of political ecosystems will consider them as:

- *Contextual.* Interactions between people, technologies, and data occur at specific settings. People are identifiable individuals or groups, geographically bounded or connected online (or both); technologies include specific devices and tools (plat-

[8] In Figure 3 we use icons from the Noun Project (https://thenounproject.com/): group icon by Gregor Cresnar; data icon by IcoDots; mobile device icon by Vildana.

forms, apps, sensors, etc.); data comprises particular datasets with different formats (open data, linked open data, etc.) and licenses of use.

- *Blended.* Interactions between people take place seamlessly, both offline and online. Global initiatives, or local initiatives that become transnational, may set local chapters where people can meet offline, organise, and discuss.
- *Distributed*. Political ecosystems are distributed networks with multiple nodes (as opposed to centralised or decentralised systems). Most likely, different political ecosystems will exhibit different connectivity maps—or political 'connectomes', to borrow an emerging concept from the neurosciences [21]. Likewise, we will need to develop and refine an appropriate 'connectomics' [20] to analyse their structural connections.
- *Open ended.* Political ecosystems will evolve and adapt as the context changes. Stakeholders and their interests are not stable, technologies change rapidly and data has been characterised with the 4 Vs (volume, velocity, variety, and veracity). In this regard, a political ecosystem can be viewed as an adaptive complex systems.
- *Technologically agnostic*. Political ecosystems rely on civic technologies that can be replaced. Whereas specific technologies can fail, be prohibited, or its supply be interrupted, there is a possibility for alternative implementations.
- *Modular*. Participation and civic engagement are fluid concepts that can adopt multiple forms. Civic technology tools now support a vast range of options for citizens and groups: data collection, fact checking, monitoring, signing petitions, crowdfunding, ideating, deliberating, drafting, voting, etc. In a modular political ecosystem, these options are available to cater for different levels of interest and engagement. Some forms of engagement will likely attract large numbers, while some others, requiring more time and cognitive effort, will appeal smaller crowds.
- *Scalable*. Political ecosystems should be able to accommodate increasing numbers of nodes (participants, technologies, data) and interactions between them without compromising connectivity and effectiveness.
- *Reusable knowledge*. Political ecosystems tap on collective intelligence to produce new forms of collective, commons-based knowledge. This knowledge may adopt multiple formats: unstructured conversation threads in forums, websites, social media, portals; annotated documents and wiki-documents, crowdsourced legislative and policy drafts, proposals, manifestos, etc.; infographics, reports, case-study repositories, podcasts, videos, etc. Both deliberation and epistemic approaches to democracy assume the need to find and reuse knowledge in deliberation and decision-making processes. Josiah Ober adds to this necessity the dimension of problem solving, in the sense that untapped knowledge can only be 'discovered' in relation to a particular political issue by making a connection of relevance between that knowledge and the issue at hand [14,15]. From a linked democracy approach, we are interested in discovering how those connections are made and how they can be reused.
- *Knowledge-archiving*. To reuse politically relevant knowledge, political ecosystems need to find ways to trace and reproduce such knowledge. Traceability, reproducibility, and accountability are essential components of collective, commons-based knowledge.

- *Aligned*. Political ecosystems may emerge bottom-up, as civic engagement initiatives, or top-down, from legislative or open government initiatives. In any case, only if institutional arrangements are in place there will be the consequential decision making and feedback loops that characterise aligned processes.

5 Concluding Remarks

In this paper we have briefly sketched some tensions and trilemmas in democratic theory that are relevant to the topic of designing civic technologies for democracy. Our contention is that technology can provide solutions to these tensions and trilemmas if we embed the issues at stake in a particular institutional meso-level.

Most online platforms focus on facilitating engagement and specially participation. As we have seen, it is not possible to scale up participation by mere technological prowess. Developing technological platforms in the near future will require an integrated approach where trade-offs between political values are explicitly acknowledged and the institutional design of the different components and processes is coherent with contextual constraints and changing environments. Civic values are also critical, and we agree with the perspective of Shannon Vallor [24] when she states that 'the designs of such platforms have assumed civic virtues as inputs, rather than helping to cultivate them—virtues like integrity, courage, empathy, perspective, benevolence, and respect for truth necessary to fuel any democratic technology, analog or digital'. A model of linked democracy is proposed to pay attention to these different dimensions.

References

1. Aitamurto, T., Galli, J.S. and Salminen, J. (2014). Self-selection in crowdsourced democracy: A bug or a feature?http://www.jorgesaldivargalli.com/w3/papers/AitamurtoSG14.pdf
2. Casanovas, P., Mendelson, D. & Poblet, M. (2017). A Linked Democracy approach to regulate health data. *Health and Technology*, doi: 10.1007/s12553-017-0191-5
3. Cormon, P. (2015). Swiss Politics for Complete Beginners (2 ed.), Geneva: Slatkine.
4. Fishkin, J. S. (2011). Deliberative democracy and constitutions. *Social Philosophy and Policy*, 28(01): 242-260.
5. Griffith, M., Spies, N.C., Krysiak, K., McMichael, J.F., Coffman, A.C., Danos, A.M., Ainscough, B.J., Ramirez, C.A., Rieke, D.T., Kujan. L., & Barnell, E.K. (2017) CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nature genetics* 1, 49(2): 170-4.
6. Kim, J., Cheng, J., & Bernstein, M.S. (2014). Ensemble: exploring complementary strengths of leaders and crowds in creative collaboration. In *Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing*; 745-755.
7. Lafont, C. (2015). Deliberation, participation, and democratic legitimacy: Should deliberative mini-publics shape public policy? *Journal of Political Philosophy*, *23*(1): 40-63.
8. Landemore, H. (2017). Inclusive constitution making and religious rights: Lessons from the Icelandic experiment. *The Journal of Politics*, *79*(3): 000-000.

9. Lloyd, A. (2017). Disentangling Democracy From Geography. *The Atlantic*. May 2, 2017, https://www.theatlantic.com/technology/archive/2017/05/disentangling-democracy-from-geography/524124/
10. Luz N., Poblet M., Silva N., & Novais P. (2015) Defining human-machine micro-task workflows for constitution making. In: Kamiński B., Kersten G., Szapiro T. (eds) *Outlooks and Insights on Group Decision and Negotiation. GDN 2015*. Lecture Notes in Business Information Processing 218: 333-344.
11. Mansbridge, J., Bohman, J., Chambers, S., Christiano, T., Fung, A., Parkinson, J., & Warren, M. E. (2012). A systemic approach to deliberative democracy. In J. Parkinson and J. Mansbridge (eds.) *Deliberative systems: deliberative democracy at the large scale*. Cambridge University Press: 1-26.
12. Meijer, A., Van der Veer, R., Faber, A, & Penning de Vries, J. (2017). Political innovation as ideal and strategy: the case of aleatoric democracy in the City of Utrecht. *Public Management Review*, 19(1): 20-36.
13. North, D. C., Wallis, J. J., & Weingast, B. R. (2009). *Violence and social orders: a conceptual framework for interpreting recorded human history*. Cambridge University Press.
14. Ober, J. (2008a). *Democracy and knowledge: Innovation and learning in classical Athens*. Princeton University Press.
15. Ober, J. (2015). *The rise and fall of classical Greece*. Princeton University Press.
16. Ng, Y. F., Coghill, K., Thornton-Smith, P., & Poblet, M. (2016). Democratic representation and the property franchise in Australian local government. *Australian Journal of Public Administration*. doi: 10.1111/1467-8500.12217.
17. Noveck, B.S. (2017). Five hacks for digital democracy. *Nature* 544**,** 287–289 (20 April 2017). doi:10.1038/544287a
18. Poblet, M., Casanovas, P., Rodriguez-Doncel, V. (forthcoming, 2017). *Linked Democracy: Foundations, tools, and applications*. Springer Open.
19. Theocharis, Y., & van Deth, J. W. (2016). The continuous expansion of citizen participation: a new taxonomy. *European Political Science Review*: 1-24.
20. Seung, S. (2013). *Connectome: How the brain's wiring makes us who we are*. Boston: Mariner Books.
21. Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS computational biology*, *1*(4), e42.
22. Urbinati, N. (2004). Condorcet's democratic theory of representative government. *European Journal of Political Theory*, 3(1), 53-75.
23. Urbinati, N., & Warren, M. E. (2008). The concept of representation in contemporary democratic theory. *Annual Review of Political Science*, 11: 387-412.
24. Vallor, S. (2017). Lessons From Isaac Asimov's Multivac, *The Atlantic* (May 2, 2017), https://www.theatlantic.com/technology/archive/2017/05/lessons-from-the-multivac/523773/

The economics of crypto-democracy

Darcy W. E. Allen[1], Chris Berg[1], Aaron M. Lane[1]✉ and Jason Potts[1]

[1] School of Economics, Finance and Marketing, RMIT University, Melbourne, Australia
✉aaron.lane@rmit.edu.au

Abstract. Democracy is an economic problem of choice constrained by transaction costs and information costs. Society must choose between competing institutional frameworks for the conduct of voting and elections. These decisions are constrained by the technologies and institutions available. Blockchains are a governance technology that reduces the costs of consensus, coordinating information, and monitoring and enforcing contracts. Blockchain could be applied to the voting and electoral process to form a crypto-democracy. Analysed through the Institutional Possibility Frontier framework, we propose that blockchain lowers disorder and dictatorship costs of the voting and electoral process. In addition to efficiency gains, this technological progress has implications for decentralised institutions of voting. One application of crypto-democracy, quadratic voting, is discussed.

Keywords: Blockchain, Cryptoeconomics, Democracy, New comparative economics, New institutional economics, Transaction cost economics, Voting

1 Democracy as an economic problem

Democracy is an economic problem insofar as it consists of a choice subject to constraints made by acting agents with diverse preferences about their own ends (Buchanan and Tullock 1962). As in market exchange, in democratic choice these constraints are transaction costs and information costs, and are determined by the prevailing institutions and technologies available to individual voters, candidates, political parties, and electoral agencies. Democratic institutions include laws governing elections and participation, rules controlling the provision of political information (such as free speech or limits to free speech, speech or donation disclosure, truth in advertising laws, or electronic advertising bans), and norms about democratic participation. Democratic technologies include those which enable the distribution of information and knowledge about democratic choice (such as the printing press or social media) and facilitate the making of democratic choice (such as printed ballot papers). Constitutionally, societies have to determine who gets to choose (the franchise), the domain over which that choice is exercised (what social choices are to be governed democratically rather than through market processes), and the mechanism by which that choice is exercised (both the form of the democracy—i.e. representative or participatory—and the electoral system—i.e. proportional or majoritarian). At a lower level, the institutional choices consist of the timing and location of elections, mechanisms to enroll and verify the identities of voters, the physical means by which the vote is made and recorded, whether individual votes are made in public or are secret, the process by which votes are counted, along with how they are verified, protected from tampering, and reported to a body for tallying.

All these decisions are constrained by the technologies and institutions available. Voter identification provides an example of a democratic institution limited by the prevailing level of technology. Before the British Reform Act of 1832, "the would-be voter appeared at the poll, tendered his vote, and then there swore an oath prescribed by statute to the effect that he had the requisite qualification" (Maitland 1908, p. 355). While the number of eligible voters was small, this was a small burden – in small boroughs individuals were likely to be recognized at the ballot box. The Reform Act both expanded the franchise and mandated the creation of an electoral roll across Britain. These procedural changes prevented disputes about eligibility occurring at the ballot box itself, but were also expected by their proponents to reduce the cost of the election (Seymour 1915, p. 107). Enrolling to vote in Australia in the twenty-first century requires either an Australian driver's license or an Australian passport—each with a color photograph of the holder and digital security features—or the verification of an existing enrolled voter how has previously passed the same.

As this suggests, technological and institutional changes have both expanded democratic possibilities and helped develop trust that individual votes—i.e. choices—are inputs into the social choice governed by the constitutional system. Technological advancement opens up alternative systems through which democracy might be practiced. Representative democracy as it stands in the twenty-first century developed world has been set according to the technological and institutional limits of prior centuries. In order to underline this point, it is worth a brief diversion into the role that technology played in equally 'democratic' but significantly different forms of democracy that have prevailed in the past.

Ancient Athenian democracy was organised predominately by sortition rather than representation. Several hundred offices, including the membership of the governing Council of the 500, were filled each year by random allotment. Athenian juries were also filled by lottery, as they still are today. For Aristotle, sortition was the defining characteristic of Athens' identification as a democracy, and, as Headlam (1891, p. 1) writes, for the modern mind 'there is no institution of ancient history which is so difficult of comprehension as that of electing officials by the lot'. Nevertheless, Athenian democracy faced many of the same practical constraints involving the selection and identification of potential office-holders and jurors. Participation in the lottery was not compulsory, but for those who chose to do so, identification was verified by ownership of a bronze identity plate. These plates were slotted into a tall marble machine, the kleroterion, from which they were withdrawn according to the random roll of a dice. Offices were allocated on the basis of the order the plates were withdrawn. The machine was introduced first to reduce possible jury tampering (Ober 1989, p. 101), and Dow (1939) suggests that the potential for fraud to be committed by the operators of the machine was prevented by running the procedure twice. Sortition was valued in part as a response to agency problems derived from political power (Berg

2015; Rancière 2009). The introduction of the kleroterion, alongside the identification controls of the bronze plates, provided a material increase in the 'democraticness' of Athenian democracy, according to that society's own conceptions of participation. In that case, technology and technological change expanded the institutional possibilities of democracy and reduced the costs of those institutions.

In this paper, we consider the same potential with blockchain technology. The next section will introduce the blockchain technology and consider its application for the institutions of voting and elections, drawing on new comparative economics and transaction cost economics to provide a theoretical framework for analysis. In the final section, we consider quadratic voting as an implication of crypto-democracy.

2 Blockchain and crypto-democracy

In 2008, Satoshi Nakamoto authored a white paper introducing blockchain technology (Nakamoto 2008). Using the complex mathematics of cryptography, blockchains enable dispersed and pseudonymous people to coordinate information and govern exchange in a decentralized way. A blockchain acts as distributed publicly accessible and secure ledger of information (Barta and Murphy 2014; Swan 2015). The first and most famous application of blockchain was through the digital currency Bitcoin (Antonopoulos 2014; Böhme et al. 2015; Godsiff 2015). This was an effort to provide a trusted non-territorial digital currency that was not reliant on a centralized bank and to operate through financial intermediaries. But the potential applications of blockchains are much broader than currency. For instance, blockchains may disintermediate and decentralize law, contracts and government (Atzori 2015; Economist 2015a; Mougayar 2016; Popper 2015; Vigna and Casey 2015; Wright and De Filippi 2015). They can facilitate self-executing smart contracts in areas such as financial derivatives and gambling (Buterin 2014; Kõlvart et al. 2016; Szabo 1997), and create distributed autonomous organizations (De Filippi and Mauro 2014). Most generally, blockchains compete with centralized hierarchical organization, such as firms and governments. Functionally this implies blockchains are a technology for creating new decentralized institutions (Davidson et al. 2016). To the extent that modern economic growth is explained through the evolution of effective institutions, blockchain may prove to be a general purpose institutional technology impacting many sectors and industries (Allen 2016; MacDonald et al. 2016).

Blockchains have also been raised as a potentially efficient solution for voting (Barnes and Brake 2016; Daniel 2015; Osgood 2016). This application has been termed 'crypto-democracy' (Davidson et al. 2016). The successful entrepreneurial application of blockchain involves outcompeting existing institutions for solving particular economic problems. Using the institutional possibility frontier (IPF) framework (developed within new comparative economics) we can compare the existing institutions for voting and the electoral process and examine the effect of the introduction of blockchain.

There is no single institution for managing the voting and election process; rather we can observe several institutional forms that exist on a spectrum of institutional possibilities. In making institutional choices society face a tradeoff between the costs of disorder, and the costs of dictatorship. How different institutions minimise these costs can be mapped as an IPF (Djankov et al. 2003). Before examining the costs of dictatorship and disorder in the electoral process, it's first important to note that these costs are subjectively perceived by each political actor (Allen and Berg 2016). Therefore, we can, for instance, use experts' perceptions of electoral integrity to understand this cost tradeoff (Norris and Grömping 2017), as well as other historical examples of social losses from the democratic process.

The costs of disorder for voting and the electoral process refer to the risk of private expropriation such as individuals committing fraudulent registration, impersonation, or voting multiple times. Prosecutions following elections provide evidence that these are more than hypothetical risks to the system (e.g. The Electoral Commission 2016). To the extent that voters have a preference in any poll, the failure of these preferences to be captured by the system—e.g. measured by voter turnout—also represent disorder costs.

The costs of dictatorship are the public expropriation of the voting process by public actors. This could include overt practices such as ballot-stuffing, vote rigging and manipulated results, which may happen where electoral officials favor the incumbent candidate or ruling party (Norris and Grömping 2017). Dictatorship costs will be present where the centrally controlled electoral register is inaccurate, either through ineligible voters being registered or eligible voters left off the list (Norris and Grömping 2017). Dictatorship costs include not just public malfeasance, but also negligence. An example of this is in the Australian 2013 Federal election, where the High Court ruled that the Senate election for the State of Western Australia was invalid because the Australian Electoral Commission had lost 1370 ballot papers (*Australian Electoral Commission v Johnston* [2014] HCA 5, 2014). Some phenomena will reflect costs of both disorder and dictatorship. One example of this is bribery, where the distinction will depend on whether it is a public or private actor that is collecting the bribe. The same can be said of integrity of the system, and the costs of enforcing the results. Violence is yet another example. That is, disorder is present when private actors deny other individuals from exercising their voting rights, such as through violence or fear of violence (e.g. Norris and Grömping 2017), whereas dictatorship will be present in instances of state-sponsored violence (e.g. Schedler 2002).

Centralised and decentralized institutions manage these dual costs in different ways. Centralised institutions limit the perceived costs of disorder by having a centrally managed voter registry and having full authority over the conduct of elections, and limits costs involved in duplication, but increases the perceived costs of dictatorship because these circumstances introduce risks that the process could be (internationally or negligently) manipulated by state actors to favor a party or candidate. Laws maintaining the electoral commission's independence guard against the worse of the perceived dictatorship costs. In contrast, decentralised institutions limit the dictatorship costs associated with concentrated power by introducing competition and choice between jurisdictions, but this introduces the risk of perceived costs of disorder by giving more power to individuals and relying on private collective action.

At this point, we can begin to construct an institutional possibilities frontier for managing the voting and election process, illustrated in Figure 1. First, on the right of the IPF, a single centralised electoral authority, controlled by the ruling candidate or party in an election. Second, a centralised electoral authority established as impartial and independent of the government of the day (e.g. the Australian Electoral Commission, responsible for conducting the electoral system for federal representatives across the country). Third, a decentralised system with several electoral authorities (e.g. in the United States, each state is responsible conducting elections of their own federal representatives). Fourth, on the left of the IPF, an arrangement of multiple privately managed systems (e.g. there are several for-profit services that provide voting and election services, used mainly by public companies and membership organisations).

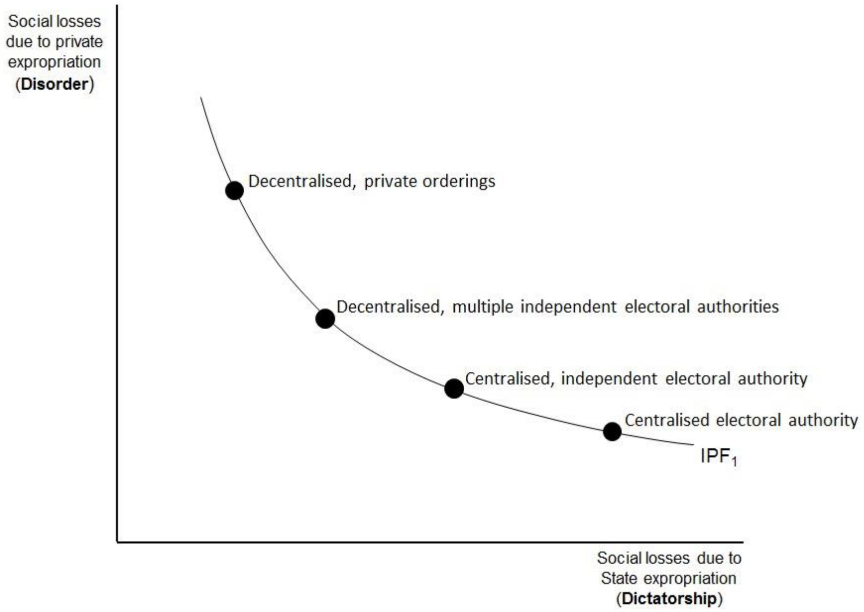

Social losses due to private expropriation (Disorder)
Decentralised, private orderings
Decentralised, multiple independent electoral authorities
Centralised, independent electoral authority
Centralised electoral authority
IPF_1
Social losses due to State expropriation (Dictatorship)

Fig. 1. Institutions of voting and the electoral process

Let us now return to the effect that blockchains have on the institutional environment. Blockchains are a governance technology reducing the costs of consensus, coordinating information, and monitoring and enforcing contracts. Indeed, given that democracy is itself an economic problem of coordinating preferences—with various potential comparatively efficient institutional solutions—it is somewhat unsurprising that blockchains may be applied to democracy. At the time of writing the most prominent application for blockchain for online voting is FollowMyVote.com, who claims to embody “all of the characteristics that a legitimate voting system requires: security, accuracy, transparency, anonymity, freedom, and fairness” using blockchain (followmyvote.com 2017). Claims over the potential of blockchain technology for voting are in effect arguing that blockchain technology comparatively decreases the various costs of dictatorship and disorder, including “robustness, anonymity and transparency” (Lee et al. 2016). Put another way, following the transaction cost economics framework of Oliver Williamson (1975), we can view blockchains as economising on the costs of uncertainty and opportunism in a decentralized way.

Of course, there is the potential that crypto-democracy could be applied within a centralised institutional possibility. A centralised electoral commission could, for example, use blockchain technology to maintain their electoral roll which has integrity and transparency benefits, meaning that the voting process would be harder to manipulate and it would reduce the possibility of human error. But we anticipate that the major benefits for crypto-democracy will be for decentralized institutional possibilities ordinarily typified by higher perceived costs of disorder, as a decentralised ledger decreases the many of those costs (e.g. fraudulent registration, security, enforcement, duplication, etc.) without needing to rely on central control. For this reason, we propose that the introduction of the blockchain technology to the voting process—crypto-democracy—causes an inward shift in the IPF, skewed towards reducing the perceived costs of disorder. This is shown in Figure 2.

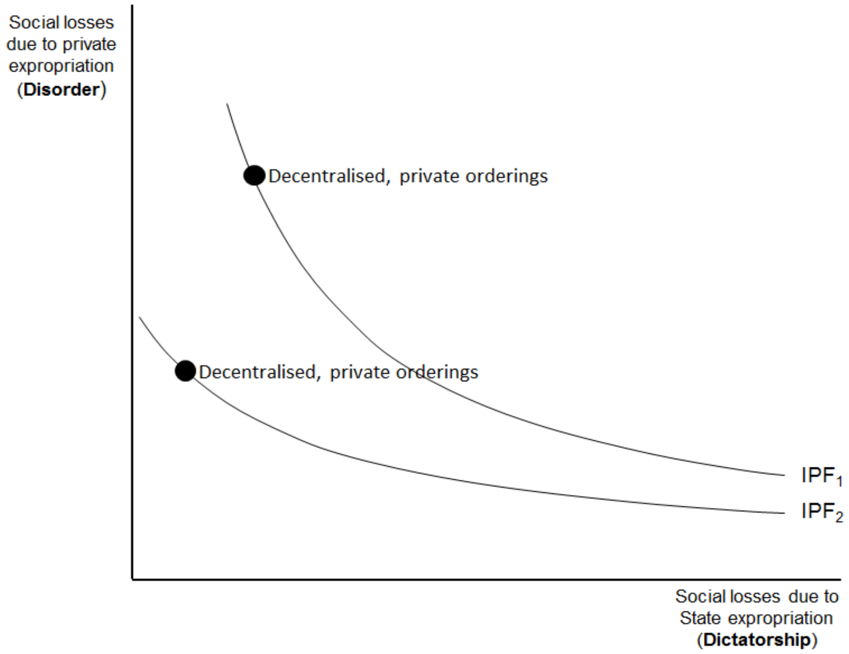


Fig. 2. Introduction of the blockchain technology

The majority of current proposals focusing on using blockchain for voting examine what appear to be pure efficiency gains for voting on the blockchain. However, an inward shift in the IPF due to the discovery of blockchain technology also presents the possibility of institutional entrepreneurship to discover new possibilities within the IPF space for solving the broader democratic problem (see Allen and Berg 2016). That is, the implication of an inward shift of the IPF implies more institutions are possible, not what those institutions are in practice. We explore one new institutional possibility to solve the democratic problem in the following section, quadratic voting.

3 A new institution of democracy: quadratic voting on the blockchain

Quadratic voting (QV) is a new voting mechanism proposed by Lalley and Weyl (2014). Posner (2016) suggests that "Quadratic voting is the most important idea for law and public policy that has emerged from economics in (at least) the last ten years". The basic idea is that the millennia old democratic franchise model of one-person-one-vote (1p1v) has the unfortunate but well-known flaw in that it is economically inefficient because it entirely ignores intensity of preference. If I care only a little about an issue and you care a lot (maybe it affects you more), we both have an identical voting margin. This leads to well-known problems with 1p1v such as tyranny of the majority. This means that issues that affect a minority of citizens, yet have significant welfare consequences for them (Lalley and Weyl offer gay marriage as an example), can be blocked by a casually indifferent majority. This is Pareto inefficient: there are clear opportunities for gains from trade. Lalley and Weyl (2014, p 2) explain that "1p1v offers no opportunity to express intensity of preference, allowing inefficient policies to persist. … The basic problem is that 1p1v rations rather than prices votes, resulting in externalities across individuals." They propose that the QV mechanism can resolve this problem (see also Posner and Weyl 2014).

QV works by introducing a payments mechanism into voting but, crucially, each voter is on both sides of the market: you pay to vote (buying votes along a quadratic pricing schedule, e.g. if 1 vote costs $1, 2 votes costs $4, 3 votes costs $9, 10 votes costs $100, 100 votes costs $10,000), but you also get paid after the vote (the payments go into a pool to be redistributed among all voters). QV is therefore both a vote pricing schedule and a reallocation mechanism. Lalley and Weyl (2014) show that the QV mechanism is, in the limit, 'robustly efficient' (Lalley and Weyl 2014, p 1) (recall the 1p1v mechanism is not efficient): QV induces revelation of true preferences, aggregates those preferences, and then compensates those affected by the decision.

There are several points to note about the QV mechanism: it overlooks persuasion; it has implementation challenges; and it has high transactions costs. First, it implements an exchange and compensation mechanism (which is the logic of seeking to improve the Pareto efficiency of an outcome where all citizens have given preferences). But an alternative mechanism—implicit in the 1p1v mechanism when understood in the context of an economy—is that citizens may seek to persuade each other to change their preferences, or to adopt better preferences.[1] The economic logic of this has recently been developed by Almudi et al. (2017) and Potts et al. (2017) in an evolutionary group selection (replicator dynamic) model they call ‘utopia competition’, in which agents use their own economic resources to seek to persuade other agents to adopt their own ‘utopia’ preference bundle. Evolutionary utopia selection model preserves 1p1v, but the compensation mechanism works through costly persuasion rather than transfer. However, the claim is that the utopia selection is also more efficient than 1p1v.

Second, as an abstract mechanism QV is asymptotically efficient. But there are still a number of implementation challenges for secure voting in relation to verifiability, robustness against false accusations, and secrecy. Park and Rivest (2016) have proposed a number of specific mechanisms using cryptographic techniques (including homomorphic encryption and zero-knowledge proofs) to resolve the issues of anonymity and payments efficiency using cryptocurrency. However, they acknowledge that the problem of overcoming collusion (which is an inherent instability in QV, which Lalley and Weyl acknowledge but offer no solution) remains problematic. However, the central message of Park and Rivest (2016) is that many of the problems of robustness in implementation can be resolved by adding cryptography to the mechanism.

A third constraint on QV, and arguably the most immediately practical problem at any non-trivial scale of application, is high transaction costs. That makes it infeasible in practice compared to 1p1v, which is for all its Pareto economic inefficiency is actually a low cost solution in exchange and contract because there is no exchange and contract (and thus has high transactions cost efficiency). This is a point that neither Lalley and Weyl (2014) nor Posner and Weyl (2014) really address. We therefore emphasise that the ‘crypto’ solution to robustness suggested by Park and Rivest (2016) also extends to a general transaction cost solution in the form of QV on the blockchain.

Quadratic voting should be understood as a mechanism that is inherently implemented on a blockchain at the point of voter identification, robustness and verification of the bidding and tallying mechanism, and security and transactional efficiency of the vote buying, fund pooling, and redistribution mechanism. By envisaging and implementing the QV mechanism in the context of a platform such as Ethereum, which enables smart contracts in which a citizen preprogram their preferences and then allow their software agent (or Distributed Autonomous Organization) to in effect automate the trades and voting and to make and receive payments, the transactions cost constraint on QV in an analog world is significantly reduced. The shift to a blockchain-platform also suggests other prospective applications that address problems of collective decision making over distributions of preference intensity, but which for

[1] This critique was also made by Tyler Cowen on his blog *Marginal Revolution*: http://marginalrevolution.com/marginalrevolution/2015/01/my-thoughts-on-quadratic-voting-and-politics-as-education.html

transactions costs reasons get caught in low Pareto efficiency mechanisms, such as the turgid representative democracy of corporate governance or city councils.[2]

4 Conclusion

The basic economic problem of democracy is to coordinate preferences between distributed people. This is an institutional problem, constrained by transaction costs and information costs, and therefore available technologies. Given that blockchain is an institutional technology for creating decentralized institutions, in this paper we have examined the potential for blockchain to open up new institutional possibilities of crypto-democracy. We focused on one new institutional possibility opened up through blockchain, quadratic voting, and its potential to more effectively solve the democratic problem.

References

1. Allen, DW and Berg, C 2016, 'Subjective Political Economy', <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2799032>.
2. Allen, DWE 2016, 'Discovering and developing the blockchain cryptoeconomy', RMIT University.
3. Almudi, I, Fatas-Villafranca, F, Izquierdo, L, Potts, J 2017, ‘Economics of utopia: A co-evolutionary model of ideas, citizenship, and socio-political change’ *Journal of Evolutionary Economics* (forthcoming).
4. Antonopoulos, AM 2014, ‘Mastering Bitcoin: unlocking digital cryptocurrencies’, O'Reilly Media Inc.
5. Atzori, M 2015, ‘Blockchain Technology and Decentralized Governance: Is the State Still Necessary?’ <http://ssrn.com/abstract=2731132>.
6. *Australian Electoral Commission v Johnston* [2014] HCA 5, 2014, <http://eresources.hcourt.gov.au/showCase/2014/HCA/5>
7. Barnes, A and Brake, CP, Thomas 2016, Digital Voting with the Use of Blockchain Technology, Plymouth University.
8. Barta, S and Murphy, RP 2014, Understanding bitcoin: a liberty lover's guide to the mechanics and economics of crypto-currencies.
9. Berg, C 2015, *Liberty, Equality & Democracy*, Connor Court Publishing, Ballarat, Victoria.
10. Böhme, R, Christin, N, Edelman, BG and Moore, T 2015, 'Bitcoin', *Journal of Economic Perspectives*, 29(2): 213-38.
11. Buchanan, JM and Tullock, G 1962, *The calculus of consent, logical foundations of constitutional democracy*, University of Michigan Press, Ann Arbor.
12. Buterin, V 2014, Ethereum White Paper: A Next Generation Smart Contract & Decentralized Application Platform, Ethereum.
13. Daniel, M 2015, 'Blockchain Technology: The Key to Secure Online Voting', *Bitcoin Magazine,* 27 June 2015.

[2] For instance what Potts et al (2017b) call *quadratic zoning* (Quadratic voting + blockchain = quadratic zoning). This weighted voting mechanism combined with redistribution of funds enables efficient coalition formation and internal transfers to create welfare maximizing urban rezoning. With quadratic zoning there is no need for urban planning, but rather urban zoning can evolve.

14. Davidson, SR, De Filippi, P and Potts, J 2016, 'Economics of Blockchain', paper presented to Public Choice, US, March 2016.
15. De Filippi, P and Mauro, R 2014, 'Ethereum: the decentralised platform that might displace today’s institutions', *Internet Policy Review*, 25.
16. Djankov, S, Glaeser, E, La Porta, R, Lopez-de-Silanes, F and Shleifer, A 2003, ‘The new comparative economics’, *Journal of Comparative Economics,* 31(4): 595-619.
17. Dow, S 1939, 'Aristotle, the Kleroteria, and the Courts', *Harvard Studies in Classical Philology*, 50: 1-34.
18. Economist, T 2015, 'The promise of the blockchain: The trust machine', The Economist, 31 October 2015.
19. followmyvote.com 2017, Blockchain Technology in Online Voting, 2017, <https://followmyvote.com/online-voting-technology/blockchain-technology/>.
20. Godsiff, P 2015, 'Bitcoin: Bubble or Blockchain', in Agent and Multi-Agent Systems: Technologies and Applications, Springer, pp. 191-203.
21. Headlam, JW 1891, 'Election by lot at Athens', Prince consort dissertation, 1890. thesis.
22. Kõlvart, M, Poola, M and Rull, A 2016, 'Smart Contracts', in The Future of Law and eTechnologies, Springer, pp. 133-47.
23. Lalley, S, Weyl EG 2014, ‘Quadratic Voting’, <https://ssrn.com/abstract=2003531>
24. Lee, K, James, JI, Ejeta, TG and Kim, H 2016, 'Electronic Voting Service Using Block-Chain', *The Journal of Digital Forensics, Security and Law: JDFSL*, 11(2): 123.
25. MacDonald, TJ, Allen, DWE and Potts, J 2016, 'Blockchains and the Boundaries of Self-Organized Economies: Predictions for the Future of Banking', in P Tasca, T Aste, L Pelizzon & N Perony (eds), Banking Beyond Banks and Money: A Guide to Banking Services in the Twenty-First Century, Springer International Publishing, Cham, pp. 279-96.
26. Maitland, FW 1908, *The Constitutional history of England: a course of lectures*, Cambridge University Press, Cambridge.
27. Mougayar, W 2016, *The Business Blockchain: Promise, Practice, and Application of the Next Internet Technology*, 1 edn, Wiley, New Jersey.
28. Nakamoto, S 2008, 'Bitcoin: A peer-to-peer electronic cash system', *Consulted*, vol. 1, no. 2012, p. 28.
29. Ober, J 1989, *Mass and elite in democratic Athens: rhetoric, ideology, and the power of the people*, Princeton University Press, Princeton, N.J.
30. Osgood, R 2016, 'The Future of Democracy: Blockchain Voting'.
31. Park, S, Rivest R 2016 ‘Toward secure quadratic voting’, <http://eprint.iacr.og/2016/400.pdf>
32. Popper, N 2015, *Digital Gold: Bitcoin and the Inside Story of the Misfits and Millionaires Trying to Reinvent Money*, Harper.
33. Posner, E. (2016) ‘Quadratic voting’ <http://ericposner.com/quadratic-voting/>
34. Posner, E, Weyl, G 2014 ‘Voting squared: quadratic voting in democratic politics’ Coase-Sandor Institute for Law and Economics. University of Chicago Law School
35. Potts, J., Almudi, I., Fatas-Villafranca, F. (2017) ‘Utopia competition: A new approach to the micro-foundations of sustainability transitions’ *J. Bioeconomics*, 19(1): 165-185.
36. Potts et al. 2017b, ‘Quadratic Zoning’.
37. Rancière, J 2009, *Hatred of democracy*, Paperback edn, Verso, London.
38. Schedler, A 2002, ‘The Menu of Manipulation’, *Journal of Democracy*, 13(2): 36-50.
39. Seymour, C 1915, *Electoral reform in England and Wales: the development and operation of the parliamentary franchise, 1832-1885,* Yale historical publications Studies, Yale University Press, New Haven.
40. Swan, M 2015, Blockchain: Blueprint for a New Economy, O'Reilly Media, Inc.

41. Szabo, N 1997, 'The idea of smart contracts', Nick Szabo's Papers and Concise Tutorials.
42. The Electoral Commission (2016), 'Analysis of cases of alleged electoral fraud in the UK in 2016' <http://www.electoralcommission.org.uk/__data/assets/pdf_file/0020/223184/Fraud-allegations-data-report-2016.pdf>
43. Vigna, P and Casey, MJ 2015, *Cryptocurrency: How Bitcoin and Cybermoney Are Over-turning the World Economic Order,* Random House.
44. Williamson, OE 1975, 'Markets and hierarchies', New York, pp. 26-30.
45. Wright, A and De Filippi, P 2015, 'Decentralized blockchain technology and the rise of lex cryptographia', <https://ssrn.com/abstract=2580664>

Promoting Public Deliberation in Low Trust Environments: Australian Use Cases

Liam Lander[1] and Nichola Cooper[2]

[1] Charles Sturt University, Melbourne, Australia
[2]Centre for the Future, Melbourne, Australia
liamlander@gmail.com
nicholacooper@protonmail.com

Abstract. A vacuum of public trust in Australia has met with the maturation of technologically competent constituents. Changing sociopolitical attitudes and perceived government corruption and inefficiency have effected demands for accountability and transparency. Two responses are visible: the digitisation of government services and original models of digital democracy. This paper discusses the role distributed ledger technology plays in decentralised governance in Australia.

Keywords: Trust, blockchain, distributed ledger, technology, democracy, open data, government.

1 Introduction

> *'A sense of the future is behind all good politics. Unless we have it, we can give nothing - either wise, or decent to the world.'* [4]

There are notable trends becoming visible to even the casual Australian observer: the widening of class structures, deepening mistrust in authority, the increasing penetration of more complex technology and living services that provide design solutions for operational or governance-related problems. The concurrent development of secure transmission architecture on accessible platforms creates a solutions environment that begins to address the primary obstacle to public engagement with authority and artefacts thus far: trust.

Increased voter cynicism, symptomatic of the politics of trust,[1] changing patterns of media consumption, the heightened exposure of political actors to public scrutiny and poor performance in economic policy, have eroded the capacity of elected representatives to govern. [20] [9] [3] In Australia, declining levels of trust are concomitant

[1] According to sociologist John Thompson, the electoral success of governments and political parties has become increasingly bound to the perceived credibility and integrity of their leaders. Within the politics of trust, campaigning on the moral failings of partisan opponents has become a strategic means of differentiating political parties in the absence of significant policy divergence based on class or ideology.

with the emergence of numerous electoral phenomena such as: poor civic engagement (particularly among young people)[2], declining political party membership[3] and reduced satisfaction with representative democracy, government, major political parties and the performance of politicians [2] [10]. This trend is not unique to Australia, of course. Trust levels are falling internationally [18], [24], however, there have been significant evolutions and variations to traditional Australian democracy therefore. These include government reform agendas, policy developments that place community trust at the centre of implementation [11], the government commission of new communication technologies and blossoming entrepreneurial endeavours into new forms of democracy.

Somewhat geographically isolated, alternative models to the traditional representative democracy have only recently touched Australian shores. Liquid, participatory, deliberative, direct and crowdsourced democracy have been designed, tested and implemented in Europe for many years now –with varying degrees of success. However, liquid democracy is relatively new to Australia. Following in the footsteps of global open-source successes such as Democracy Earth,[4] vTaiwan,[5] (g0v),[6] Pol.is,[7]

[2] A 2004 study commissioned by the Australian Electoral Commission has revealed that only one in four young people perceive politicians can be honest and fewer than half believe that politicians can be trusted to do what is right for the country [53].

[3] As few as one percent of Australia's adult population are registered members of political parties, mirroring similar declines throughout European democracies [3].

[4] Democracy Earth (http://democracy.earth/#about) is building an open source and decentralised democratic governance protocol called Sovereign backed by Y-Combinator. Their open source platform held a pilot during the Columbian referendum, allowing ex-patriots to vote when the government decided not to reopen voter registration during the referendum. Crucially, appealing to the importance of the liquid democracy model, Sovereign allowed voters to both delegate votes and vote separately on specific parts of the referendum. Instead of absolute approval or rejection, the majority of Columbians voted yes to the referendum, but no to allowing the guerrillas to participate politically. This option was a nuance that the vote, which rejected the peace deal, lacked [16].

[5] vTaiwan (virtual Taiwan) is a direct consequence of the Sunflower student demonstration demanding the rejection of a Beijing trade deal, legislation permitting the monitoring of Chinese agreements and citizen conferences discussing constitutional amendment. They use Pol.is distribute social media adverts and broadcast a public meeting where scholars and officials respond to issues that emerged in the conversation. This is followed by an in-person stakeholder meeting co-facilitated by civil society and the government, and broadcast to remote participants for the Government to bind its action to consensus, or provide a detailed explanation of why those consensus points are not (yet) feasible.

[6] G0v.tw (http://g0v.tw/en-US/about.html) is an online community that focuses on information transparency, freedom of speech and open data. They publish open-source code and develop information platforms and tools for citizens to participate in society. G0v 'rethinks the role that government plays in a digital native generation'. They believe transparency of information can help citizens better understand how government works to understand issues faster so they can hold government accountable and deepen the quality of democracy.

[7] Pol.is (https://pol.is) is a mobile platform that uses AI and machine learning to build tools that offering transparency through decentralisation and insights.

Democracy OS,[8] and D-Cent[9] (who launched Finland's Open Ministry and Iceland's Better Reykjavik programme), Australia is beginning to host their own entrepreneurs, designing the future of liquid democracy.

2 The Future of Democracy

Futurists and technologists have been alert for changes to democracy since Alvin Toffler wrote in *Future Shock* (1970) that representative government was the political technology of the industrial era. It was Toffler's vision that the electorate would be sufficiently proactively informed of likely outcomes of prospective policy to be engaged in strategic decision-making. This future-oriented, participatory, approach to policy design could have such political impact as to 'be the salvation of representative politics - a system now in crisis' [21]. Toffler's editor, Clem Bezold, wrote in *Anticipatory Democracy* in 1978 that cyber democracy would be comprised of: cyber administration, cyber voting, cyber participation, cyber agenda-setting and cyber infrastructure. In drawing attention to Australia's emerging actors in participatory democracy, this paper will briefly discuss existing and evolving global platforms that enable the automated administration of executive function, the engagement of policy makers and crowdsourcing of legislation as well as cloud and distributed ledger digital voting platforms.

Australia began to see Toffler's vision realised in 2016. Not formally - in amendments to the traditional federal electoral format; the Australian Electoral Commission is bound by the regulations of the Electoral Act, 1918. Instead, alternative direct methods of voting are becoming known with rising numbers of minor parties such as Vote Flux[10] and the Pirate Party (Australia)[11] as well as movements such as Online Direct Democracy (Senator Online, of old),[12] MiVote,[13] The Fourth Group[14] and or-

[8] DemocracyOS (http://democracyos.org/) is an online space for deliberation and voting on political proposals, using software that aims to stimulate better arguments and come to better ruling through peer collaboration. It is a platform for a open and participatory government.

[9] D-CENT (Decentralised Citizens ENgagement Technologies, https://dcentproject.eu) was a Europe-wide project joining citizen-led organisations that have transformed democracy in the past years, and helping them develop the next generation of open source, distributed, and privacy-aware tools for direct democracy and economic empowerment. The EU-funded project started in October 2013 and ended in May 2016. D-CENT tools inform and deliver real-time notifications about issues that matter, they propose and draft solutions and policy collaboratively; decide and vote on solutions and collective municipal budgeting; and finally implement and reward people with blockchain reward schemes. The tools can be combined in ways to support democratic processes.

[10] www.voteflux.org

[11] The Pirate Party campaigns for a free society, civil liberty, and trust in the rule of law. They believe in the right to privacy and transparency in government and organisations. Pirate Party Australia was founded by Rodney Serkowski in 2008 and has grown from a small group of activists to a fledgling political party.

[12] Online Direct Democracy (https://www.onlinedirectdemocracy.org/) is a not-for-profit, entering candidates for the upcoming federal election. They claim to be Australia's first internet-based registered political party aiming to provide everyone listed on the electoral roll with a

ganisations such as Our Say.[15] Advancements in technology have made it possible for these groups to cost-effectively overcome barriers to entry - each designing for trust proactively or iteratively, using such technology as the blockchain to overcome lows in public confidence and initiate participatory forms of democracy.

Current data by Edelman suggests that public confidence in government functioning and satisfaction with democracy is so low as to pose a challenge to the legitimacy of government [5]. Long perceived as perpetuating a culture of cronyism and corrupt behaviour, Australians have gradually invested less trust in their elected political representatives with only one in four Australians now believing politicians can be trusted [14]. The most recent report of the Australian Election Study indicated only 60% of Australians were satisfied with democracy and only 12% of the population believed the nation was governed with the interests of all Australians in mind [2].

A similar sentiment is represented in many Western democracies, where public confidence in political leadership and representative democracy has steadily eroded since the 1970s. Attributed to social, political, technological and economic factors associated with globalisation, contemporary neo-liberal political outcomes and the changing distribution of labour, a concurrent belief that the system is failing is raising individual and community fear, exciting the rise of populist parties and movements [5]. Existing political trust research has examined the execution of civic responsibility as a function of trust [23] and found that civic participation does affect trust in two pathways: ethical behaviours and service competence. Ethical behaviour is defined as operating when officials transcend self-interest or agency priorities to pursue public needs and service competence is defined as an ability to develop goods and services that achieve sustained public satisfaction. Findings in Wang and Van Wart's study suggest that the public trusts the administration more when demand and response for services is well met during the participation process, and the public perceives a high level of satisfaction with the services provided. This met need results in greater horizontal trust, driving participation in civic duty that results in greater vertical trust [23]. First, service must be delivered to the public's standards and ethically.

Results of Australian empirical studies [7] suggest the cause of democratic entropy in Australia is increasingly ascribed to the performance and behaviour of political officials and division between representative democracy and participatory democracy functions reinforcing a national culture of anti-politics. Findings by Evans *et al.* demonstrate that if politicians support participatory politics with the objective of reinforcing the function of representative democracy to ultimately develop a more integrated, inclusive and responsive democratic system, Australians may trust and engage

direct voice in parliament. Once elected, Online Direct Democracy MPs are bound by their agreement with the party to act on behalf of their constituents and all Australians.

[13] www.mivote.com.au

[14] www.thefourthgroup.org

[15] Founded in Melbourne in 2010, Our Say is a collaborative platform that connects community leaders with members of the public. Designed to build trust and authenticity in public communication and decision making it has been used by high profile politicians such as Julia Gillard who describe the platform as "...modern democracy and modern technology at work" ((https://www.oursay.org).

more with the process of democracy [7]. Accordingly, this paper discusses the use of nascent technologies, such as distributed ledger technology, and the potential impact on public trust in democracy through the case studies of MiVote and Vote Flux; two fledgling Australian direct democracy start-ups, operating on blockchain platforms.

3 The Potential Blockchain Offers Democracy

The blockchain underpins distributed ledger technology; the first use case for which was Bitcoin. It operates in a decentralised peer-to-peer network using cryptographic algorithms to verify, validate and distribute transactions across millions of nodes, enabling the secure, auditable, transmission of assets without intervention by central authority. I.e., the function of decentralised trust (or trust-by-computation) facilitates the automation of instructions (also known as smart contracts), which may obviate the role of third parties and reduce administration and management costs.

Since, theoretically, anything of value can be stored on the distributed ledger: contracts, certifications, music, art, identities, policies, bills and votes, for example, governments are beginning to invest in blockchain for improved efficiencies and performance in regulatory compliance, contract and identity management and civic services [12].

Concurrently, we see increasing numbers of use cases both designed with the intent to mediate distrust by instilling transparency into process and circumvent trust entirely by disintermediating the relationship between voter and representative (or consumer and supplier). Top-down applications include movements towards open government and the prevalence of open-data and the bottom-up use of blockchain technology to store and transmit data securely.

Notwithstanding policy makers' concessions to the public's call for transparency, the mechanics of government have remained largely unchanged since federation. Empirical studies have hitherto indicated poorly designed or implemented democratic innovations risk greater mistrust and the Australian government is acutely aware of, and commercially sensitive to, mistrust 'choking the use and value of Australia's data…' To that end, the government believes 'improving trust community-wide is a key objective'. [11] The Australian Government's report *Ahead of the Game—the 2010 blueprint for the reform of the Australian Public Service (APS)*—is cast in this light.

Pertinent to the argument in favour of blockchain technology's application to centralised services is the consensus algorithm that is fundamental to achieving trust. Unlike traditional human service-related transactions, such as depositing money, sending a parcel or achieving settlement on a property, where we trust an unknown person to conduct the transaction with integrity; a blockchain transaction does not require trust. An information sender does not need to trust the peer network; valid transactions are automatically verified (or rejected) on confirmation of the appropriate cryptographic code (proof-of-work) and further distributed throughout the network until the transaction has reached every node on the network. The proof-of-work algorithm requires miners to resolve a time-consuming and complex mathematical puzzle for the network nodes to achieve consensus and deem the transaction reliable.

This consensus model of governance reduces the risk of fraud; enables the automated processing of smart contracts, creates economies and efficiencies and the open network instills trust in the transparency and auditability of ledgers. This shift from trusting people to trusting maths offers numerous opportunities for blockchain technology in low-trust environments.

For governments, distributed ledger technology is an ideal infrastructure for the digital storage or publication of central records and smart contracts permit the reliable administration of routine government functions. As part of the Delaware Blockchain Initiative, the Governor of the State of Delaware asked the state's Bar Association to consider clarifying Delaware corporate law to authorise, track and transfer shares on a distributed ledger. The first milestone on the Initiative's rollout plan has passed at the Delaware Public Archives; using smart contracts to automate compliance with laws regarding the retention and destruction of archival documents. The second milestone is to be completed in late 2017: smart Universal Commercial Code (UCC) filings. The current process is paper-based, slow and error-prone, UCC filings on a distributed ledger automate the release and renewal of UCC filings, reduce mistakes, fraud and cut cost.

In local civic functions blockchain technology could be applied to the democratic process to increase trust and engagement given the public perception of governments as "somewhat of an encumbrance – too slow, too corrupt, too lacking in innovation, and benefiting too few". [1] However, state and federal blockchain-based electoral voting is considered too novel for Australian application.

Further to a call for submissions into e-voting by the Victorian Parliament's Electoral Matters Committee in 2017, the committee decided in favour of electronic ballot paper scanning at the 2018 Victorian state election despite Australia Post's submission of a blockchain voting architecture. For, the committee were not satisfied that the interconnectivity between government and citizenry was foolproof; citing technology failures experienced during the Census website crash and Centrelink data hacking as a 'salutory lesson'.[16] In addition to security, a significant contributor to the VEC's decision to preference electronic ballot paper scanning was cost. The cost per vote for the vVote electronic voting system at the 2014 Victorian state election was $2,261.85 per vote (gross). Excluding capital implementation costs, the cost of a

[16] The committee received 34 submissions from organisations including, but not limited to, Australia Post, Elections ACT, Electoral Commission Queensland, Electoral Council of Australia and New Zealand, Australian Electoral Commission, Tasmanian Electoral Commission, New South Wales (NSW) Premier and Cabinet, the Research School of Computer Science, the Australian National University and Computing and Information Systems department at the University of Melbourne. The committee were informed of the risks associated with e-voting in lower security and verifiability of the NSW iVote and Victorian vVote system compared to the scrutineering of paper ballots; the technology especially vulnerability to a 'man in the middle' attack. Accordingly, working with the Australian Electoral Commission (AEC) the committee recommended '...an Electronic Voting Board oversee scrutiny…' of the '…most rigorous security standards available...'. The committee were not, however, satisfied the interconnectivity between government and citizenry was foolproof [15].

vVote at the 2014 Victorian state election was $396.46, the New South Wales iVote system cost approximately $9.50 per vote, and around $10.60 at the 2015 NSW state election [15]. In contrast, excluding capital costs, a blockchain vote costs approximately $1.

At a grassroots level, there are multiple examples of community efforts in designing innovative models of democracy and democratic processes that have been tested in Scandinavia,[17] Europe and the United States of America[18]- on distributed ledger technology and cloud platforms. While some projects have failed to achieve social scale, governments have adopted some; Iceland's Your Priorities and Spain's Decide Madrid were the result of community collaboration following the 2008 global financial crisis. New models of democracy are not only the result of crisis, however, but declining trust in politicians and democratic institutions. A political donations crisis preceded the inception of the Estonian Citizens' Assembly, the work of the President and civil society organisations that ultimately proposed democratic reform.[19] Innovation in Australian democracy could similarly be attributed. The following case studies offered in MiVote and Vote Flux are the consequence of dissatisfaction with political financing, perceived corruption and the influence national and international political donors have in the formation of public policy.

4 The Future of Australian Democracy

Toffler worried humans were racing blindly into the future without reflection or consultation. His vision for the future of democracy was inclusive; imagining that the public could more effectively steer legislation: "We need to, quite literally, go to the people with a question that is almost never asked of them: What kind of a world do you want ten, twenty or thirty years from now? We need to initiate, in short, a continuing plebiscite on the future...backed with technical staff to provide data on the social and economic costs of goals, the trade-offs so that participants may make reasonably informed choices among alternative futures...not merely expressed as vaguely expressed, disjointed hopes, but coherent statements of priorities for tomorrow" [21].

This vision is realised in our case studies, MiVote and Vote Flux, which use blockchain to invite consultation on the formulation of policy, and in the founding princi-

[17] Direktdemokraterna is a Swedish party that uses cloud-based voting for referenda in a liquid democracy model. https://direktdemokraterna.se/hur-ska-det-ga-till/

[18] Collaborative and co-design approaches have been applied to democratic decision-making on e-democracy platforms such as Germany's Adhocracy (https://adhocracy.de/), America's Challenge.gov (https://www.challenge.gov/list/), Decide Madrid (https://decide.madrid.es/?locale=en), Estonia's Rahvaalgatus (https://rahvaalgatus.ee/), Iceland's Your Priorities (https://yrpri.org/domain/3). Some of these tools are gaining traction: Your Priorities has been used in Romania, the UK and Estonia. Decide Madrid is being used by municipal governments in Barcelona, A Coruña and Oviedo.

[19] The Estonian Citizens' Assembly Process (2013) was the direct result of a legitimacy crisis involving Estonian political parties and representative institutions caused by illegal political financing. Government responded using democratic innovation: eliciting public support in crowdsourcing and deliberative mini-publics.

ples of Online Direct Democracy (ODD). All three Australian organisations are united in motivation: engendering the inclusive participation in non-partisan politics free of influence. Like Toffler, these organisations believe the constituency contains the inherent skills and wisdom necessary to make ethical and appropriate choices for the benefit of their community but they use technology to bridge the divide between the constituency and representatives. Their approaches are broadly similar: inform the public of tabled issues before parliament and the consequences of the bill, seek the opinion of the constituency and feed this information directly to Flux, MiVote or ODD parliamentary representative to vote in accordance with the majority opinion. There are fundamental differences, however.

Online Direct Democracy is a registered political party that crowdfunded and built PollyWeb as a secure voting platform on similar security principles as banking systems, with three-step authentication. Their platform enables Australians to discuss, rate and vote on bills and amendments as they are tabled in parliament. PollyWeb engages the public in political dialogue by undertaking research into tabled issues before parliament, providing relevant resources and then polling the public on their opinion regarding the issue. This opinion poll is then communicated to the ODD party representative to consider in their vote. ODD ran two candidates in the 2016 federal election and received 11,133 votes or 0.09% with the highest vote achieved in the state of Queensland with 0.23% of the total votes going to the party.[20]

4.1 Flux

The classical definition of democracy is an idealised principle of government whereby the rule of society is derived from the popular will of the people [14]. Vote Flux was founded in 2015 and they operate a custom Issue Based Direct Democracy (IBDD) model founded on *Deutschian Fallibilism*, an evolution of Popperian Fallibilism and David Deutsch's book *The Beginning of Infinity*. IBDD preferences problem-solving over representing "the will of the people."[21] Their policy position evolves as a consequence of a voting auction market where a neutral central liquidity token allows voters to move their political capital to issues of most immediate subjective importance. In forcing an opportunity cost to voter choice, IBDD interrupts 'tyranny of the majority' in the search for good policy;[22] achieved by the trading of votes to subject matter experts.[23]

[20] https://www.onlinedirectdemocracy.org/

[21] https://voteflux.org/2017/05/26/an-overview-of-flux-and-ibdd/

[22] https://voteflux.org/2017/05/26/an-overview-of-flux-and-ibdd/

[23] In practice, each Flux member receives one vote for each bill before parliament. This vote may be traded for a credit in the case of low interest issues or conserved for a later vote of greater interest. Additional liquidity tokens can be collected and distributed for issues voters consider of particular importance that are designed to be inflationary in value. Thus, a more contested piece of legislation will cost more to gain more votes; a less contested piece of legislation extra votes will cost less. In so doing, IBDD seeks to engage apathetic constituents that may otherwise waste their vote in the representative system, by providing a mechanism to trade their vote with someone more knowledgeable or energised by the outcome of the issue.

Vote Flux is a registered political party with 6269 members (as at 12/7/17 but are growing at an average growth rate of 30.4% per month) and branches in each state. They ran candidates in the 2017 Western Australian state elections, unsuccessfully. Co-founded by a software developer, the Flux application is designed on their SecureVote blockchain platform, which can support in excess of 1 million votes a minute, or 1.5 billion votes in 24 hours. Using a private audit log an independent third party can verify a personal identity against a blockchain identity but a patented two-step process of "oblivious shuffle" means no one else will be able to link the two. This ensures each vote comes from an anonymised registered voter [25].

4.2 MiVote

MiVote employs a model of destinational democracy - almost precisely as Toffler imagined in 1970 [21]: "...a continuing plebiscite on the future…". With founding principles of neutrality, transparency, representation and equality, their approach is inclusive and participatory in nature. After rigorous research of a pertinent issue, four strategic directions are applied for the constituency to consider and vote on. Written accessibly, with basic, intermediate and advanced cascading levels of information, the research serves to inform the public of the facts and impact of the issue and asks them how they would prefer their representatives vote on their behalf.[24]

MiVote is a movement with 2765 members - it is not yet a registered political party. Currently, their blockchain voting platform consults the membership base gathering data points regarding sentiment. Their intent is not to run in state elections but to propose candidates for the next federal election, using the platform as a direct communication between the voter and their MiVote representative. The objective is to direct parliamentary action in favour of the majority opinion.

5 Limitations: Why Change Will Be Slow

The relationship between citizen and state hereafter may be shaped by the influence of emerging technology but this will not be strictly limited to the blockchain. Advancements in distributed ledger technology and machine learning will disintermediate processes on ever more grand scales at the grassroots level, growth in the use of cloud-based platforms are encouraging collaboration and internal hacking of government processes indicate democracy in Australia is changing - distributed ledger technology is only one indicator of which.

[24] This might be represented, for example, as reform made to the Political Donations Bill, framed as: increased public funding, removal of public funding, donations made to candidates or no change to the bill at all - maintenance of the status quo. MiVote's ranking system, similar to the Single Transferable Vote, means constituents vote for what is most acceptable. Their consent-based decision making approach is reinforced by intermittent polling of the constituents, enquiring of issues most important to them; this forms part of the research agenda.

Sociopolitical behaviour in Australia indicates favourable responses to participatory platforms. Evans, Halupka and Stoker found in their 2016 study that investments made into projects that would enhance trust in the political system and elected representatives would be well received. Their primary finding included justification for a national democratic audit to answer three questions: how do Australians imagine their ideal democracy? What do they expect from politicians within it? How is the present system failing? [7].

The increasing number of social organisations in Australia that provide tools and strategies to increase citizen engagement, political participation and trust is testament to this. There are at least twenty-five organisations undertaking deliberative decision-making or process design making deep strides into reforming public engagement at a community and structural level [8].

Accordingly, we find two trends that will influence the expression of Australian democracy that mirror European precedent: the integration of open-source participatory platforms by government agencies that promote transparency and encourage public trust and the exponential growth of secure, decentralised platforms that attract early adopters to digital democracy. The following reasons indicate why blockchain technology is unlikely to be a feature of government's participatory platforms:

- ***Blockchain is slow***: continued development in open-source distributed platforms such as Ethereum, Omni Layer,[25] the lightning network,[26] and Hyperledger[27] already suggest the imminent faster processing of data and more scalable databases. Increasing numbers of interoperability protocols and off-chain transactions will also eventually obviate performance concerns. For, the fundamental limitation to faster adoption is directly tied to the primary benefit of blockchain technology: the trade-off made between security and speed. The process of data mining means that blockchain cannot deliver speed and security simultaneously without compromising on the number of nodes on the network. Vote Flux may have their permissioned blockchain network finalised in time for the next federal election, which would advance the processing of votes from 1-3 per second to millions per minute, but this may cause public criticism with regards security.
- ***Distributed ledger technology is new*****:** until rigorous testing of a novel technology has proved consistently reliable by international governments it is improbable we will see the adoption of distributed ledger technology for large-scale government functions in the short-term. This means the proving ground for liquid democracy models in Australia is the start-up enterprise and minor political parties.
- **Scale:** the novelty of the technology means there is currently limited available empirical data and academic studies in wide-ranging implementation and achieving social change; this titrates investment which impacts product awareness and viability. As demonstrated in Europe and with changing funding approaches by Flux, MiVote and ODD, it is famously difficult to achieve social scale within resource

[25] http://www.omnilayer.org/
[26] The Lightning Network: https://lightning.network/
[27] https://www.hyperledger.org/

allocation for civic technology organisations. Unless organisations are inclined to partner and share resources there are risks of reduced impact and public weariness.

- ***The matter of the digital divide*****:** creators of blockchain-enabled democracy platforms are regularly asked about accessibility. If representative democracy is progressively becoming elitist how does introducing novel technology designed on premium platforms reduce this? Social research into political participation identifies that the deeper the vein of socio-economic inequality and more prevalent the social complaints, the more people participate in the political process [2]. To encourage participation and social cohesion, platforms need to be considered as accessible as possible or we compromise political equality and fracturing democracy into a greater number of off-shoots.

6 Conclusion

Society's most historic structures are undergoing challenge by the equalising, unrelenting forces of technology and globalisation. This paper described the two primary responses by governments and entrepreneurs: the publication of open data to increase transparency and public trust and the use of blockchain technology to disintermediate the mistrusted process.

Using Alvin Toffler's prescient vision of an inclusive, consultative society utilising a participatory democracy model, we briefly discussed three Australian organisations realising this vision. Two of which are using distributed ledger technology to defend against the primary criticism e-voting has endured so far: security. While the Australian government is reticent to apply untested technology to federal functions it is researching the implications of blockchain, as are nine in ten governments [12].

Per Evans, Halupka & Stoker's findings [7], supported by politicians, a combination of cloud-based and decentralised technologies that support the public in engaging with participatory decision-making may ultimately enable society to reorganise around principles of horizontal trust, enhancing social capital and decreasing class stratification; but this is a long-term view. What is clear from the research is that technology is not a panacea for increasing public engagement or trust. A multi-faceted response is required that engages with community action groups, technologists, civil technology firms and industry to design bespoke engagement mechanisms until more direct alternatives are deemed suitable.

References

1. Atzori, M. (2015). Blockchain technology and decentralised governance, is the state still necessary? University College of London - Center for Blockchain Technologies. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2709713
2. Cameron, S.M. and McAllister, I, Trends in Australian political opinion: Results from the Australian election study 1987–2016, ada.edu.au/.../Trends%20in%20Australian%20Political%20Opinion%201987-2016.pdf
3. Clarke, H. D., Sanders, D., Stewart, M. C., & Whiteley, P. Political choice in Britain. Oxford: Oxford University Press (2004)

4. C.P. Quoted in Bezold, C. Anticipatory Democracy. Random House: Toronto, Canada, p.105 (1978)
5. Edelman, Trust barometer, http://www.edelman.com/
6. Eisenstadt, S, N.. The Order-Maintaining and Order-Transforming Dimensions of Culture. In Eisenstadt, S, N. Power, Trust and Meaning: Essays in Sociological Theory and Analysis. The University of Chicago Press, Chicago. P. 306–327 (1995)
7. Evans, M., Halupka, M & Stoker, G. (2016). Who do you trust to run the country? Democracy, trust and politics in Australia. University of Canberra. Retrieved from http://www.governanceinstitute.edu.au/magma/media/upload/media/943_Who-do-you-trust-to-run-the-country.pdf"
8. PoliVote, http://polivote.com.au/product/oz-nz-spirit-of-democracy/
9. Hetherington, M, J. The Political Relevance of Political Trust, American Political Science Review, 94(4), 791-808 (1998)
10. Hill, L., & Rutledge-Prior, S., Young People and Intentional Informal Voting in Australia, Australian Journal of Political Science, 51(3), 400-417 (2016)
11. Holmes, B. Citizens' engagement in policymaking and the design of public services, Parliament of Australia (2011)
12. IBM Institute for Business Value, https://www-935.ibm.com/services/us/gbs/thoughtleadership/blockchain-for-government/
13. Karlson, M., Jonsson, M. & Åström, J., Did the Estonian Citizens' assembly help restore political legitimacy? Analyzing changes in vertical and horizontal trust among participants Paper prepared for presentation at the ECPR General Conference in Montreal (2015)
14. McAllister, I. Keeping them Honest: Public and elite Perceptions of Ethical Conduct Among Australian Legislators, J. Political Studies, 48(1), 22-37 (2000)
15. Parliament of Victoria. Inquiry into electronic voting, Electoral Matters Committee. (2017)
16. Fast Company, https://www.fastcompany.com/3068382/can-technology-save-democracy
17. Rolls, E., Emotion and Decision-Making Explained. Oxford Scholarship Online (2013)
18. Ruscio, M. & Kenneth, P., Trust, Democracy and Public Management. J. Public Administration Research and Theory 8 (3), 461-477 (1996)
19. The Guardian, https://www.theguardian.com/world/2015/jun/12/madrid-manuela-carmena-deal-socialists-mayor
20. Thompson, J, B., Political Scandal: Power and Visibility in the Media Age. Cambridge: Polity Press (2000)
21. Toffler, A. Future shock. Pan Books: Great Britain, London, p.436 (1970).
22. Warren, M, E. (Eds). Democracy and Trust. Cambridge University Press, New York (1999)
23. Wang, X. & Van Wart, M. When public participation in administration leads to trust: an empirical assessment of managers' perceptions. Public administration review (2007)
24. Whitley, P., Clarke, H. D., Sanders, S., & Stewart, M. Who do Voters Lose Trust in Governments? Public Perceptions of Government Honesty and Trustworthiness in Britain 2000-2013, The British Journal of Politics and International Relations, 18(1), 234-254 (2016)

Intelligent Warning Systems: ‘Nudges’ as a Form of User Control for Internet of Things Data Collection and Use

Rachelle Bosua[1], Karin Clark [2], Megan Richardson [2] and Jeb Webb[1]

[1] School of Computing and Information Systems, University of Melbourne, Australia
[2] Melbourne Law School, University of Melbourne, Australia
{rachelle.bosua,karin.clark,m.richardson,jeb.webb}@unimelb.edu.au

Abstract. The modern digital world of networking and connectivity makes possible a new era of computing in which users exert greater control over the collection and use of their personal data through the Internet of Things (IoT). Our recent empirical work indicates that traditional forms of consent are inadequate and that users are looking for different levels of and greater involvement in controlling the collection and use of their personal data – with some participants voicing particular concerns about collection and use of sensitive data, such as health information, and others pointing to particular risks, such as insecure storage in the Cloud. In response to these needs we propose a new *Intelligent Warning Application* in the form of a conceptual architecture for an App that empowers users to control their IoT data collection through users: 1) identifying their own levels of risk, 2) customizing the App allowing for the setting of their identified risk levels, and 3) situated use of the App warning users of risk-averse situations through ‘nudges’. We conclude with a discussion illustrating scenarios of the App’s.

Keywords: Internet of Things, Privacy, Data Protection, Intelligent Warning Systems, Nudges.

1 Introduction

The uncontrolled collection of user data through the Internet of Things (IoT) is becoming a matter of particular concern in a world of more connectivity, networking and collaboration afforded by the IoT. How can individuals better control the collection and use of their personal data in an increasingly connected and digitized world of the IoT? This world enables multiple new services and the exchange of information, which promise to significantly ease and enrich our lives in many different ways. For example, the flick of a single switch can instantaneously operationalize multiple devices, invoke different devices’ services and feed back information based on unique predetermined individual needs. However the mass collection, integration and use of individuals’ private and personal data through modern data mining techniques and big data algorithms lead to growing privacy concerns in Australia and around the world regarding individual privacy and protection of personal data, as well as information security [1-3].

Although the open Internet-based infrastructure on which the IoT is based facilitates tremendous access to information, there are specific features that affect data protection, privacy and security [4-6]. Firstly, 'interaction' in the context of the IoT comprises data collection between multiple machines and embedded sensors without human intervention, immediate reception or control of any personal data [6]. Secondly, entities, organizations or individuals other than individual users whose data is being collected, are in control of the data being collected through IoT devices. Thirdly, without their knowledge or consent, individuals can be followed through surveillance, while their data from different data sets can be combined and processed intelligently to infer new insights based on an individual's patterns of behavior [7]. And finally, at least in Australia, there is as yet, no legal framework that responds effectively to the diverse problems that can arise in the collection, use of and protection of individuals' privacy in the context of the IoT [1-2] [7-8].

A recent Australian study on individual attitudes towards privacy, conducted by the Office of the Australian Information Commissioner, indicates that the Australian public's concern for online privacy has increased over the last five years [3]. However, despite this expressed concern, many survey participants described apparently contradictory habitual behaviors, such as not reading privacy policies (65%) or accepting default settings while using social media (50%) instead of adjusting these settings to limit who has access to their personal information. While this may seem to lay blame at the feet of the users for choosing not to engage in the learning or administrative tasks required for assuring their personal privacy, an obvious rejoinder to this argument is that these requirements may be unreasonable given the context of modern digital communications.

Modern digital communications a) present complex processes simplistically through heuristic interfaces that hide most of this complexity from the user, and b) rely on the fact that users have been conditioned to accept personal disempowerment while using the internet. The former condition extends beyond using graphical user interfaces to spare users having to deal with programming code: actual audiences, relationships between entities, and information flows (to include who is doing what with data) are all effectively hidden from the average user or internet-connected services. The latter condition is self-evident insofar as users are routinely presented with situations that have been engineered by other parties: programs that work in certain ways and allow some forms of interaction while disallowing others. In other words, while people can engage in navigational and interactive behavior within the online environment, they often do so with limited insight or control over the implications of these behaviors. Furthermore, the providers of services typically have interests in data collection that lead them to actively obscure their interests or the details of how data is used within their business models. Conditioning users to accept situations that serve these interests can also clearly be beneficial to the provider.

In view of the above challenges and the fact that data collection through the IoT is on the increase with limited to no practical control currently exercised over individual data collection and use of this data, the question is how and to what extent individuals can be provided more support in controlling their interaction in a world of data collection enabled by the IoT? In response to this question we propose a conceptual architectural model of an Intelligent Warning Application (App) that allows users to exert more control over the collection and use of their individual data collected through the IoT. Reliant as it is on cooperation from IoT producers in providing information requested by the Intelligent Warning App, we offer this as an example of the principle of 'privacy by design' that we advocated in earlier papers based on this research [1] and [2], and which numerous privacy regulators have also endorsed. Further, based as it is on extensive user interviews and two focus groups, our proposed conceptual model also embeds a broader idea of responsive regulation i.e. regulation scaled to achieve effective regulation in response to a perceived need and with minimal intervention in preference to heavy-handed top down regulation.

This paper consists of six sections. Section 2 provides background literature that illustrates the current gap in the literature with respect to intelligent warning applications/tools. Section 3 provides background to the initial conception of our proposed Intelligent Warning App and introduces the research approach to be followed to design this tool. Section 4 presents evidence of specific user concerns and recommendations leading to a conceptual model of one view of the Intelligent Warning App design. Section 5 discusses three scenarios that illustrate instances of nudging based on a user's profile built from knowledge garnered about the user's privacy needs. The Conclusion (Section 6) elaborates on next stages of the study with some limitations and recommendations for further research.

2 Background literature

The notion of uncontrolled data collection and use of user's data is a problem that plagues many individuals in a modern world of greater connectivity and exchange of data through the Internet of Things (IoT). While the Internet is one of the most disruptive technologies of the modern age, the constant collection of data through multiple connected devices is a significant concern, especially to individuals who value their privacy. Of particular concern is the notion of *transparency* and *understanding* about how and where IoT data is collected, how IoT data is stored, and when this data is used and integrated with other data sets.

2.1 Privacy, data protection, and security in the context of IoT data collection and use

Privacy' may be treated as a broad concept covering multiple aspects of the collection and use of personal information, along with other things (for instance [9]). Alternatively, some especially European commentators may label this 'data protection' (for instance [10]), while reserving the label 'privacy' for the more particular problem of being made subject to an unwanted public gaze [11-12]. In our previous published papers we pointed out that our interviewees tended to adopt the latter view although they also considered data protection to be a pressing concern, both for them personally and also for society [1-2]. As such, this paper is concerned both with questions of privacy and data protection and (unless otherwise specified) we treat these as overlapping and congruent concerns.

It is not just the control of personal information that is at stake here. Concerns may also extend to 'security', a term that concentrates on the protection of collected information from unwanted external access, for instance from hacking. Security principles (confidentiality, integrity and availability of information) [13] guarantee that access to collected information is restricted, open only to those who are authorized to do so, and that stored information is trustworthy and accurate. The heterogeneous nature of the IoT in combination with its wide scale of use is expected to increase security risks of the current Internet. More specifically, the limited computing power of IoT technologies violates traditional security countermeasures and enforcement calling for the need to define valid IoT security and trust models to gain full acceptance by its user base [14].

2.2 Informed consent in the context of technological artefacts

One shortcoming of the IoT is the limited support offered for the exercise of informed consent i.e. giving users the ability to concur with data collection and use techniques. Specifically, the collection of personal data through the IoT, is more than often unencrypted, uncontrolled through sensors embedded in the environment, or in the form of wearables or surveillance devices concealed in the environment.

In regard to this, the desire of consumers to exert control over their data has experienced a major shift over the last two decades. While a minority concern in the 80's, by the 2000s individual fears about the potential abuse of personal (consumer) information have become a major concern [15]. Consumers have become concerned about the ways in which their personal information is both collected and used, with one study indicating that almost 88% of US Internet users have expressed their wishes to have an 'opt-in' privacy policy (in 2001) whereby Internet companies need to first obtain users' permission to share their personal information with others [15]. As a result the notion of a minimal informed consent has evolved through political, legal, economic, social and technological realms.

Informed consent has been introduced as a mechanism to gain more user trust by articulating business practices for collecting and using personal information and giving users autonomous choice in terms of data collection and use. In this regard the model of informed consent for Information Systems has been introduced in 2000 [16] constituting values associated with being 'informed' (including disclosure and comprehension) and giving 'consent' (i.e. voluntariness, competence and agreement). This model has since inception been incorporated in the 'Value-Sensitive Design' framework touted by many authors [17-21] as an integral part of large-scale real world software systems. Value sensitive designs appreciate human values in a principled [18] and comprehensive way throughout the process of designing technological artefacts.

While informed consent is an attribute of many of today's modern web-based Apps and technology artefacts, ethical considerations related to providing and substantiating informed consent is considered in a modern world of technology to be inadequate, outdated and limited [22]. More specifically, there are concerns that data collection and use practices are not clearly communicated in a responsible way to pave the way for informed consent [23]. In addition, current privacy policies are not clear and understandable by ordinary consumers in conveying how the collection and use of individuals' personal information can be protected. This problem is exacerbated in an interconnected world of the IoT.

In a world of higher levels of service delivery, enabled and facilitated by increased collection and use of personal IoT data, the notion of informed consent is therefore a major concern. Indeed, prior studies indicate that users often unknowingly 'consent' to data collection and use practices of online Apps in exchange for services, while anecdotes from our empirical research indicate that the inclusion of value-sensitive design frameworks in Internet applications as a form of gaining consent is often ignored or bypassed [1]. With the increasing collection and use of individuals' personal information, there is therefore a need for users to be more cognizant of IoT data collection and use allowing them to control these activities in a more systematic way.

2.3 Nudges as a form of control

Over the last few years the notion of 'nudges' as a form of leading or guiding individuals in certain directions while also preserving their freedom of choice, has been debated significantly (see, for instance [24-27]). Nudges as a 'soft reminder' prompting users of unacceptable online behavior have been applied in different contexts e.g to support smokers to persevere in quitting smoking, and more recently as part of the Facebook web interface nudging users to more carefully consider the content and audience of their online disclosures [28]. Being a reminder, nudges can also serve as a warning or intervention that can support users in making decisions to disclose relevant or more or less information. The notion of 'reminders' is not new, and originated as computer-based 'reminder systems' in the 90's, specifically in the context of ambulatory preventative care systems. In the medical domain reminder systems serve as invaluable prompts to alert medical staff to necessary interventions associated with treatment practices to enhance patient safety [29].

Over time the use of computer-based reminder systems has become more mainstream as evident from their use in the form of 'nudges' in other application areas such as appointment reminder systems associated with email, audit and feedback reminders systems, costs of borrowing and workflow systems that are associated with rule-based processing of information. Reminders or recommendations that are in the form of nudges and specialized forms of nudges have emerged as a form of changing behavior. This form of behavioral changing has attracted considerable attention, often leading to concrete reforms in specific domains. Nudges exist in many different forms such as the sounding of alarms that call for human intervention (e.g. in the medical domain), or reminders in the form of animations or prompts that encourage online system users to interact though the entering of data or specific input device activity. Another form of nudging encourages users to pause and reflect prior to entering or posting information/content online (as in the case of Facebook [28]). Depending on the extent of intervention required, more interactive forms of nudging could be in the form of online intelligent assistants that provide users 'intelligent guidance or warnings' calling for (perhaps guiding) specific user actions or behavior.

While there are deeper questions to be asked about nudges, for instance "what they signify and express for individuals and their capacity for autonomous and responsible decision-making" [26], the use of appropriate individual-centric and intelligence-based forms of 'nudging' may be instrumental in guiding users to exert more control over the collection and use of their personal information through the IoT, as proposed in the next section.

2.4 Towards intelligent warning systems

Based on our initial study of individual perceptions of privacy and concerns about control over IoT data collection and use [1-2], there is a need to design appropriate tools that enhance or supersede traditional forms of informed consent. In our follow-up focus groups conducted this year we were told that 'warnings' may be more useful to IoT users than further refining contract terms (especially where these are treated as non-negotiable). The incorporation of 'nudges' allowing users to define and select different levels of and forms of control over the legitimate collection and use of their IoT data is an attractive option. We therefore propose an Intelligent Warning App that complements IoT data collection by allowing individuals to exert control over the collection and use of their personal data through the IoT. To justify the development of such an App, we report on empirical work conducted to elicit requirements from users in this regard.

3 Research methodology and findings

3.1 Research methodology

As precursor to defining the functional requirements of our Intelligent Warning App, it is worth noting our research methodology. We followed an intense requirements elicitation phase to get a deeper understanding of IoT data collection and use practices and problems. Our overall aim was to gain specific knowledge of the issues from a group of IoT users and software engineers involved in the development of IoT software. We were specifically interested in concerns about privacy, data protection and security and wanted to hear the views of both sets of stakeholders to verify whether the identified problems can be tackled.

Following ethics approval, the first stage of our study comprised 24 interviews with 14 IoT users and 10 IoT designers/software engineers in October 2015 to January 2016. Interviews were individual one-hour face-to-face interviews conducted in Melbourne with IoT users and experienced software engineers in the 28 to 55 year age group. One of the authors conducted the interviews and transcribed the audio-recorded interview data, followed by an analysis of this data to identify key functional requirements. Three of the authors were involved in the data analysis to ensure triangulation and agreement of the key themes that emerged from the data. We reported on this study in two published papers [1] and [2], where we argued that laws needed to provide responsive regulation of IoT privacy/data practices, including through the encouragement of minimal standards of transparency and control integrated into the design of IoT, adopting a principle of privacy by design (a principle which is partially but by no means perfectly expressed through APP 1 of the Australian Privacy Principles under the Privacy Act 1988 (Cth), which states that APP entities should "take such steps as are reasonable in the circumstances to implement practices, procedures and systems" to ensure compliance with the APPs).

Our second stage involving 2 focus groups with 4 and 7 (total 11) users and 6 IoT designers/software engineers followed in April 2017. The aim at this stage was to confirm the veracity of the findings of our first stage before moving on to obtain a more refined understanding of user requirements for privacy, data protection and security of IoT devices and compare these with options that designers thought were feasible. Both focus groups were conducted on one day (one in the morning and the other in the afternoon), each lasting one and a half hours. All four authors were present with two authors leading the focus groups and two authors acting as observers. Focus group conversations were audio-recorded and used to confirm the key themes in the form of functional requirements outlined in the next section. Four participants in our first stage participated in the stage 2 focus groups, while the other focus group participants were new, selected on the basis of their knowledge of/interest in privacy, data protection and security related to the data practices of the IoT.

3.2 Findings

As in the case of stage 1, a number of users said that they would like to have more transparency and control over their information, as one participant stated "from the perspective of a user you don't actually know what data is collected by these devices concerning you and your habits.... cheaper, faster and smarter often means unregulated". Users also agreed that there might be individual and cultural variations in terms of what information was considered particularly sensitive and how it should be treated.

At the same time, users questioned the value of the standard term consent regimes that IoT systems typically employ, describing these as "extremely lengthy and full of legal jargon that a user does not understand" and essentially 'click, click, click' regimes that allowed little scope for negotiation or individual variance. In particular one of the participants indicated that this 'regime' is a result of "...the design of the user interface and having been trained as a user – that is the user experience to click-click and don't worry about the rest of it", adding that "there is no actual conscious thought in the process".

Instead, a number of users expressed a preference for more targeted 'warnings' that would cater to particular concerns about the level of the protection and security accorded to their information and would allow them choices as to how to respond. One software engineer indicated that "I talk about notification, about different actions you take within the software system. If a software engineer designs notifications into what are the side effects [of data collection] of whichever action I have taken within the software, it will help give users awareness about the implications of what you [the data collector] are doing".

With these in mind, the next sections of this paper focus on how such warning systems might be designed and integrated into IoT devices from an architectural view, as well as how legal standards, for instance in Australia, might be drawn on by policymakers to encourage and regulate such design features to ensure they operate to enhance rather than constrain individual capacities for autonomous and responsible decision-making.

4 Conceptual architecture of the intelligent warning app prototype

Figure 1 proposes a conceptual architectural model illustrating examples of information flows resulting from IoT data collection that the Intelligent Warning App should inform the user about. This diagram illustrates three dataflow scenarios that will nudge the user for a form of intervention depending on users' set-up preferences in respect of their data.

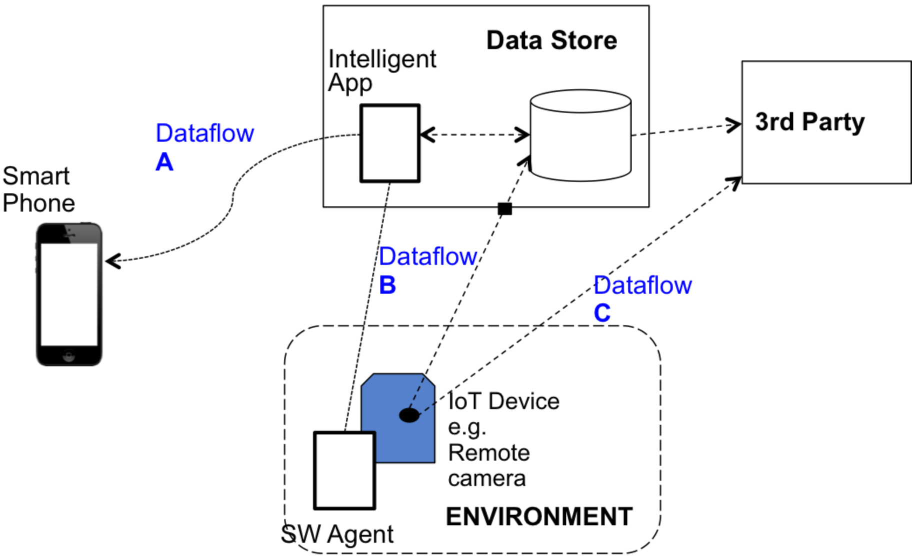


Fig. 1. Conceptual architectural model for the intelligent warning app

The above diagram represents a client server model with an IoT device and the Intelligent Warning App's intelligent agent (IA) that learns about a user's privacy and data protection requirements as set up by the user. Initially users set up their preferred protection levels, for example, control settings for i) GPS location; ii) images and iii) data movement/transfer. An initial period of use may lead to modification of the setting-level knowledge stored by the Intelligent Warning App's intelligent agent. The intelligent agent is also linked to one or more sniffers (e.g. in Figure 1 represented by one small black rectangle), which monitor traffic flows in a connected network with the consent and cooperation of the IoT service provider (who may treat this as a way of offering an externalised system of privacy-by-design to users and complying with any relevant legal obligations, for instance in the Australian case under the Privacy Act's APPs – including APP 1, noted above). The next section describes three different 'nudging' scenarios that the Intelligent Warning App will typically alert to the user.

4.1 Description of dataflow examples that will nudge the user to either consent or request adaptation of control

- Dataflow A: as set up by 'Abigail', the Intelligent Warning App will sense or track that Abigail's fitness-monitoring IoT device which is connected to her smart phone, accesses her geolocation location data through the phone's geolocation technology and integrates this data with her fitness IoT data in order to target localized advertising about health and fitness services (in a way that if not consented to, may breach a local privacy or data protection law, for instance in the Australian case APP 2: regarding a use of sensitive health information that is not 'directly related' to the primary purpose for which the information was collected, and APP 7: direct marketing using sensitive health information). Based on controls set up in the intelligent agent by the user, this activity will either inform the user or alert the user to possible actions that include closing the port through which the geolocation data flows.
- Dataflow B: as set up by 'Beatrice', the Intelligent Warning App will assess whether images or videos of Beatrice that are collected by her security camera are encrypted prior to storing these on the server. The checking of encryption is not limited to images and videos but can also be applied to any other type of data which is being sent via one or more channels from an IoT device to a server. Users will be aware through nudging that collected data is not encrypted as this data is sent out of a specific environmental boundary (in a way that again may breach a local privacy or data protection law, for instance in the Australian case APP 11 which imposes an obligation on APP entities to 'take such steps as are reasonable in the circumstances' to protect personal information that they hold from misuse, interference, loss, unauthorized access, modification or disclosure). Once again the user can decide to take preventative actions to stop the flow of unencrypted data, for instance disconnecting the device or putting the device behind a "firewall".
- Dataflow C: in this scenario, as set up by 'Chester', the Intelligent Warning App makes Chester aware of voluminous data flowing through one or more channels to a third party server overseas (in a way that again may, if done without consent, breach a local privacy or data protection laws, for instance in the Australian case APP 8 which imposes strict standards on cross-border disclosures of personal information). Once again the Intelligent Warning App will sense or track uncontrolled movement. Hence the Intelligent Warning App should 'learn' of destina-

tions of data and by knowing this, and the setting of user controls, nudge the user of any uncontrolled movement of data through specific communi-cation channels. The user might then formally act on this by consenting or reporting inadequate behavior to an appropriate regulatory entity institution (in the Australian case, the OAIC).

5 Discussion

Our recommended architecture is considered as an initial attempt to address the gaps in individually controlled data/information collection and use through the IoT. We consider the illustrated conceptual architecture in Figure 1 as the first stage towards developing a fully functional version of our proposed Intelligent Warning App. We aim to further refine our conceptual architecture into a detailed architectural design to build a prototype of the App. The next stage of this research is therefore the capturing of more detailed requirements to identify a complete and consistent set of functional and non-functional requirements to build the App and its core intelligent agent component. More specifically, the finer details of the Intelligent Warning App's intelligent agent needs to be identified to formulate detailed design requirements of its architecture in ways that will both utilize features of machine-learning effectively, and at the same time, comply with the basic legal requirements of privacy and data protection laws in multiple jurisdictions as well as broader community norms regarding the treatment of personal data, building these safeguards into the system, see [30]. We expect that a comprehensive set of semantic processing algorithms using artificial intelligence pattern matching techniques have to be designed as the core functionality of this depends on its intelligent agent component.

6 Conclusion

Our research in progress proposes one view of an Intelligent Warning App that draws on user-selected control levels and privacy principles that are aligned with Australia's Privacy Act APPs to nudge users to better control the collection and use of their private data through the IoT. We consider the model and dataflow scenarios presented here the first in a series of models (e.g. process, domain classes, service performance and use case models) that need to be developed to illustrate different architectural views of the Intelligent Warning App. Once these models are developed, a prototype App will be designed for evaluation.

This research is limited as it is in the early stages of conceptual design and prototype development and can only proceed once all functional and non-functional requirements have been defined. An Agile SDLC development approach in combination with intelligent agent-based software design is proposed for the App development. Another limitation is that the actual form of nudging as a means for users to control the flow of their data is at this stage unspecified. User-specific requirements need to be elicited through further interviews and discussions with our focus group members while the more nuanced aspects of the Intelligent Warning App's design also needs to be developed in much more depth.

References

[1] Richardson M, Bosua R, Clark K, Webb J, Maynard S and Ahmad A. (2017). Towards responsive regulation of the Internet of Things: Australian perspectives, *Internet Policy Review: Journal on Internet regulation*, 6(1).

[2] Richardson M, Bosua R, Clark H, Webb J, with Ahmad A and Maynard S (2016). Privacy and the Internet of Things, *Media, Law & Arts Review*, 21(2), pp. 336-351.

[3] Office of the Australian Information Commissioner (2017). Australian Community to Privacy Attitudes report (online resources: accessed on 20 May 2017: http://www.opengovasia.com/articles/7599-australian-community-attitudes-to-privacy-survey-shows-58-of-australians-trust-state-and-federal-government-departments

[4] Babar S, Mahalle P, Stango A, Prasad N, and Prasad R (2010). Proposed security model and threat taxonomy for the Internet of Things (IoT). In: *International Conference on Network Security and Applications* (pp. 420-429). Springer Berlin Heidelberg.

[5] Kozlov D, Veijalainen J, and Ali Y (2012). Security and privacy threats in IoT architectures. In *Proceedings of the 7th International Conference on Body Area Networks* (pp. 256-262). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[6] Weber RH (2010). Internet of Things–New security and privacy challenge. *Computer Law & Security review*, 26(1), pp 23-30.

[7] Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A and Khan SU (2015). The rise of 'big data' on Cloud computing: review and open research issues *Information Systems*, 47, pp 98-115.

[8] Weber RH (2009). Internet of things–Need for a new legal environment? *Computer law & Security Review*, 25(6), pp. 522-527.

[9] Westin AF (1967). Privacy and freedom. Atheneum New York.

[10] De Hert P and Gutwirth S (2009). Data protection in the case law of Strasbourg and Luxemburg: Constitutionalisation in action. In: *Reinventing data protection?* pp. 3-44. Springer Netherlands.

[11] Gavison R (1980). Privacy and the Limits of Law. *The Yale Law Journal*, 89(3), 421-471.

[12] Austin L (2003). Privacy and the Question of Technology. *Law and Philosophy*, 22(2), 119-166.

[13] Whitman ME and Mattord HJ (2011). *Principles of information security*. Cengage Learning.

[14] Sicari S, Rizzardi A, Grieco LA and Coen-Porisini (2015). Security, privacy and trust in Internet of Things: The road ahead. *Computer Networks*, 76, pp 146-164.

[15] Friedman B, Lin P and Miller JK (2005). Informed consent by design. *Security and Usability*, (2001), 503-530.

[16] Friedman D (2000). Privacy and technology. *Social Philosophy and Policy*, 17(02), 186-212.

[17] Friedman B and Kahn Jr PH (2003). Human values, ethics, and design. *The human-computer interaction handbook*, 1177-1201.

[18] Friedman B and Nissenbaum H (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3), 330-347.

[19] Hagman J, Hendrickson A and Whitty A (2003). What's in a barcode? Informed consent and machine scannable driver licenses. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems* (pp. 912-913). ACM.

[20] Nissenbaum H (1998). Protecting privacy in an information age: The problem of privacy in public. *Law and philosophy*, *17*(5), 559-596.

[21] Friedman B, Kahn PH, Borning A and Huldtgren A (2013). CH: Early engagement and new technologies: Opening up the Laboratory, Vol 16 of the series Philosophy of Engineering and Technology (pp 55-95) Title: Value Sensitive Design and Information Systems.

[22] Rhodes SD, Bowie DA and Hergenrather KC (2003). Collecting behavioural data using the World Wide Web: considerations for researchers. *Journal of Epidemiology and Community Health*, 57(1), 68-73.

[23] Pollach I (2005). A Typology of Communicative Strategies in Online Privacy Policies, Journal of Business Ethics, 62, pp 221.

[24] Thaler R and Sunstein C (2008). Nudge: The gentle power of choice architecture. *New Haven, Conn.,Yale*.

[25] Yeung K (2012). Nudge as Fudge. *Modern Law Review*, 75(1), 122-148.

[26] Yeung K (2017). 'Hypernudge': Big Data as a Mode of Regulation by Design. *Information, Communication & Society* 20(1), 118-136.

[27] Baldwin R (2014). From regulation to behaviour change: Giving nudge the third degree. *The Modern Law Review*, 77(6), 831-857.

[28] Wang Y, Leon PG, Acquisti A, Cranor L, Forget Al and Sadeh N. (2014). A field trial of privacy nudges for Facebook. *CHI 2014,* April 26-May 01, Toronto Canada.

[29] Meddings J, Rogers MAM, Macy M and Saint S (2010). Systems Review and Meta-Analysis: Reminder systems to reduce catheter associated urinary tract infections and urinary catheter use in hospitalized patients, Clinical Infectious Diseases, 51(5), pp 550-560

[30] Agrafioti F (2015). Privacy by Design is Key to the Future of Artificial Intelligence. *Huffington Post*, 26 October, 2015.

Linked Democracy 3.0: Global Machine Translated Legislation in the Age of AI

Sean Golz[1]

[1] University of Waikato, Hamilton, New Zealand
sean.goltz@waikato.ac.nz[1]

Abstract. This paper outlines the efforts made by Global-Regulation, a world legislation search engine, to engage artificial intelligence in two ways: (i) employing machine translation to translate the world's legislation to English and, (ii) creating an automated system to identify compliance clauses and extract penalties from legislation. This paper describes Global-Regulation's vision and technology in the context of linked democracy and the democratization of artificial intelligence.

Keywords: Machine translation, compliance, penalties, linked democracy, democratization, AI.

1 Introduction

There is a strong relationship between democracy and transparency [1] At the same time, some argue that big data will enable citizens to be governed by a data-empowered "wise king", who would be able to produce desired economic and social outcomes almost as if with a digital magic wand. [2] These trends bring to the front a term recently used by Microsoft's CEO Satya Nadella, of democratizing AI. By making AI available to everyone, it can move from a centralized tool, to one which can be used in fields such as healthcare, education, manufacturing, retail and more. The ultimate aim, is sharing AI's power with the masses, allowing anyone and everyone to use the AI systems they need [3].

This paper outlines the efforts made by Global-Regulation (www.global-regulation.com), an online search engine of the world's legislation, to engage artificial intelligence in two ways: (i) employing machine translation to translate the world's legislation to English and, (ii) creating an automated system to identify compliance clauses and extract penalties from this legislation. These means are intended to foster democracy and improve regulation by enabling lessons drawing from one jurisdiction to the other.[4] As stated by Lloyd: "The internet collapses geography and expands our concept of community, yet geographic community is a cornerstone of our structures for democratic participation"[5].

[1] The author is a co-founder of Global-Regulation.

The paper starts with a background section underpinning the problem that Global-Regulation came to solve, how the project was born, preliminary steps that were taken and the underlying motivation that drove this project's co-founders to build it. The second section deals with the method in which artificial intelligence was used in engaging machine translation on a massive scale and in the creation of the PenaltyAI system, designed to identify compliance clauses and extract penalties from the legislation. Finally, it concludes with what have been learned and what can and should be done next.

2 Background

According to Monson, "government services can and should be delivered as efficiently and effectively as the technology you use to get a ride or order dinner"[6]. In a nutshell, Monson is capturing both the problem and the underlying motivation behind Global-Regulation. Before meeting by chance, both co-founders were running novel legal websites. Addison Cameron-Huff was running a website that tracks and provide alerts on new provincial legislation and the author was running a website that provides summaries of case studies in regulation. Joining forces enabled the co-founders to embark on this ambitious project creating a search engine of the world's legislation. The co-founders did not realized at the time the scale and magnitude of this project nor the challenges lying ahead.

Very quickly it dawned on the co-founders that when one is offering a source that was never available before (e.g., Canadian academics can now read Italian legislation in English) one needs to convince users that it is valuable. More than once we heard American or Australian regulators pondering why should they look at regulations from Denmark, for example? 'We are looking only at comparable jurisdictions', they told us. 'What lessons can we possibly draw from remote parts of the world?!'. It was academics that immediately realized the potential and started exploring the database with enthusiasm. And surprisingly enough, it was the tech giants Microsoft, Google and Amazon, that generously supported our vision.

Global-Regulation is now the largest search engine of legislation from around the world, enabling comparative search of 1.6 million laws and regulations from 88 countries. Global-Regulation has employed Microsoft and Google's machine translation on a massive scale translating 750,000 laws and regulations from 26 languages into English. By providing this information, Global-Regulation unlocks the global village vision in law by automating database translation. To support its vision, it employs cloud based technology powered by Amazon to gather, index and standardize legislation from different countries across the globe. The challenge Global-Regulation faced was twofold: how to deal with laws in different languages that are coming from different legal systems.

3 Machine Translation

Initially, Global-Regulation connected to each country's official government website and uploaded the legislation to its database. This process did not enable the inclusion of legislation in foreign languages. Dealing with this challenge has bearing not only on commercial aspects, but also on Global-Regulation's founders' vision, to have the entire world's legislation searchable, in English, in one place.

The importance of this vision cannot be overstated, mainly for developing economies with unique regulatory structure interested in drawing in external investment on the one hand, and making their legal system transparent to its citizens, on the other. Making legislation transparent, accessible and searchable, especially on a comparative basis, is one of the cornerstones of democracy and a task made possible on this scale only due to recently mature technology, advances in artificial intelligence, and governments making laws available online.

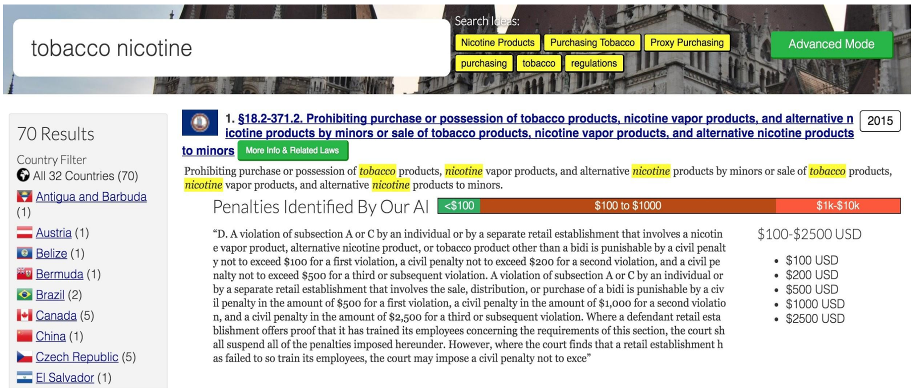
The process of machine translation for laws is as follows:

1. Index the laws in the original language and track which language the law is in (in some countries laws are published in several languages)
2. Download the laws in the original language.
3. Convert laws to "plaintext" (from HTML, XML, PDFs, etc.), where plaintext means UTF-8-encoded plain text files.
4. Format the plaintext so that items like headers, footers, and extra non-legal information is removed. Attempt to normalize line endings (especially important for PDF conversions which have odd formatting issues).
5. Break the plaintext into pieces that can be handled by machine translation systems (which generally have a size limit) using logical break points such as line endings. Also translate the title of the law.
6. Convert each piece into English then stitch the English version together using the breakpoint identified in the previous step.
7. Store the translated law and the original law in the database.
8. As machine translation models for languages improve, periodically re-translate the laws and store them in the database.

4 PenaltyAI

Following the use of machine translation, Global-Regulation have taken a step further in order to use its huge database of world laws along with the advanced capabilities of artificial intelligence. This step involved the development of a system (called 'PenaltyAI') to identify compliance clauses in legislation and extract the actual penalties from these clauses, and convert it to US dollars (if needed). This ambition to create the ultimate risk and compliance system came into life when Global-Regulation's founders realized that penalties are the kind of information that can be identified with a high degree of certainty by an artificially intelligent system [7].

After seemingly endless testing, experimenting, coding, consulting[2] and hard work, Global-Regulation presented its Search[3] – the first and only AI system that identify compliance clauses in legislation on a global scale, extracts the actual penalties amount and serves it all to the user in US dollars. See two examples in Figure 1 and Figure 2.



1. §18.2-371.2. Prohibiting purchase or possession of tobacco products, nicotine vapor products, and alternative nicotine products by minors or sale of tobacco products, nicotine vapor products, and alternative nicotine products to minors

Prohibiting purchase or possession of *tobacco* products, *nicotine* vapor products, and alternative *nicotine* products by minors or sale of *tobacco* products, *nicotine* vapor products, and alternative *nicotine* products to minors.

"D. A violation of subsection A or C by an individual or by a separate retail establishment that involves a nicotin
e vapor product, alternative nicotine product, or tobacco product other than a bidi is punishable by a civil penalt
y not to exceed $100 for a first violation, a civil penalty not to exceed $200 for a second violation, and a civil pe
nalty not to exceed $500 for a third or subsequent violation. A violation of subsection A or C by an individual or
by a separate retail establishment that involves the sale, distribution, or purchase of a bidi is punishable by a civ
il penalty in the amount of $500 for a first violation, a civil penalty in the amount of $1,000 for a second violatio
n, and a civil penalty in the amount of $2,500 for a third or subsequent violation. Where a defendant retail esta
blishment offers proof that it has trained its employees concerning the requirements of this section, the court sh
all suspend all of the penalties imposed hereunder. However, where the court finds that a retail establishment h
as failed to so train its employees, the court may impose a civil penalty not to exce"

$100-$2500 USD

- $100 USD
- $200 USD
- $500 USD
- $1000 USD
- $2500 USD

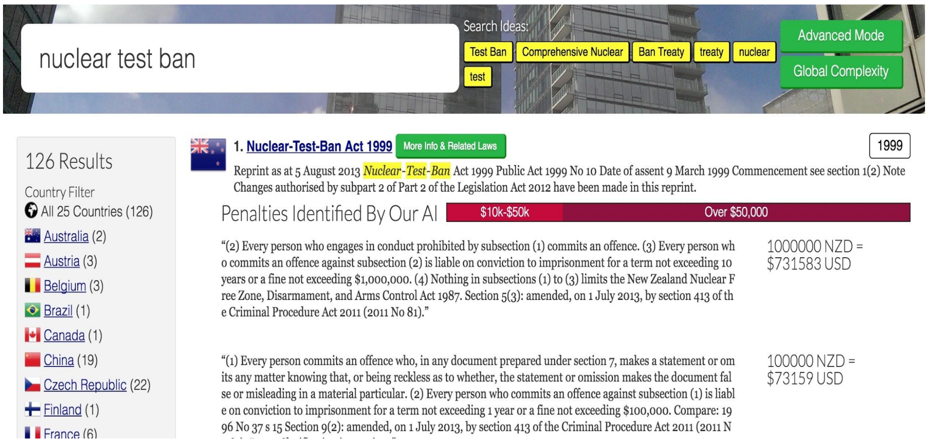
Fig. 1. PenaltyAI search results for 'tobacco nicotine' (Source: www.Global-Regulation.com)

This approach offers advantages at several levels:

- Provides an AI system that can read legal text and produce useful meaning;
- Enables risk and compliance professionals to explore real and relevant data on a global scale, in English;
- Allows governments and businesses to assess and enhance their compliance efforts;
- For researchers, it assists in comparing and contrasting risk and compliance data globally;
- Perhaps most importantly, it is a first step in enabling the public to have a transparent and informed access to regulatory compliance hence an enhancement of democracy.

[2] Thanks to Kyle Gorman from Google for the words to numbers converter recommendation.

[3] https://www.global-regulation.com/penalty_ai.php



1. Nuclear-Test-Ban Act 1999

Reprint as at 5 August 2013 *Nuclear*-*Test*-*Ban* Act 1999 Public Act 1999 No 10 Date of assent 9 March 1999 Commencement see section 1(2) Note Changes authorised by subpart 2 of Part 2 of the Legislation Act 2012 have been made in this reprint.

"(2) Every person who engages in conduct prohibited by subsection (1) commits an offence. (3) Every person wh
o commits an offence against subsection (2) is liable on conviction to imprisonment for a term not exceeding 10
years or a fine not exceeding $1,000,000. (4) Nothing in subsections (1) to (3) limits the New Zealand Nuclear F
ree Zone, Disarmament, and Arms Control Act 1987. Section 5(3): amended, on 1 July 2013, by section 413 of th
e Criminal Procedure Act 2011 (2011 No 81)."

1000000 NZD = $731583 USD

"(1) Every person commits an offence who, in any document prepared under section 7, makes a statement or om
its any matter knowing that, or being reckless as to whether, the statement or omission makes the document fal
se or misleading in a material particular. (2) Every person who commits an offence against subsection (1) is liabl
e on conviction to imprisonment for a term not exceeding 1 year or a fine not exceeding $100,000. Compare: 19
96 No 37 s 15 Section 9(2): amended, on 1 July 2013, by section 413 of the Criminal Procedure Act 2011 (2011 N

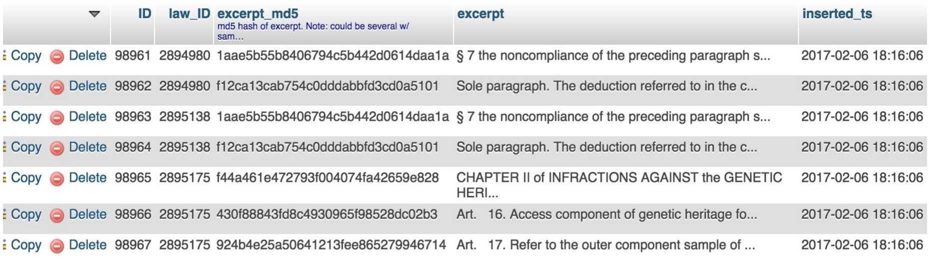
100000 NZD = $73159 USD

Fig. 2. PenaltyAI search results for 'nuclear test ban' (Source: www.Global-Regulation.com)

The "PenaltyAI Search" can answer questions like "What would I pay for violating money laundering laws in Jamaica?" or "How much would a smuggler who warehouses stolen goods in China pay if they're caught?"

The penalties are extracted by an offline algorithm that runs on an Azure Virtual Machine that performs the following steps:

1. Find laws that mention keywords associated with civil penalties (as a first pass);
2. Convert all word numbers (like "one million") into international number format ("1,000,000.00");
3. Identify the paragraphs that likely contain civil penalties based on words and numbers;
4. Merge several penalties into one, whether they related to the same "clause" (section) of a law;
5. Extract all the clauses and penalties;
6. Exclude certain classes of text that are almost never penalties but look like penalties (such as laws about gold coins and section references in laws that have to do with money);
7. Recognize currencies in text, and combine this data with our table of national currencies, and convert penalties into USD using Yahoo! Finance rates (through the XML API call);
8. Store the penalties and clauses in a MySQL database (RDS) (see Fig. 3).



▼	ID	law_ID	excerpt_md5 md5 hash of excerpt. Note: could be several w/ sam...	excerpt	inserted_ts
Copy Delete	98961	2894980	1aae5b55b8406794c5b442d0614daa1a	§ 7 the noncompliance of the preceding paragraph s...	2017-02-06 18:16:06
Copy Delete	98962	2894980	f12ca13cab754c0dddabbfd3cd0a5101	Sole paragraph. The deduction referred to in the c...	2017-02-06 18:16:06
Copy Delete	98963	2895138	1aae5b55b8406794c5b442d0614daa1a	§ 7 the noncompliance of the preceding paragraph s...	2017-02-06 18:16:06
Copy Delete	98964	2895138	f12ca13cab754c0dddabbfd3cd0a5101	Sole paragraph. The deduction referred to in the c...	2017-02-06 18:16:06
Copy Delete	98965	2895175	f44a461e472793f004074fa42659e828	CHAPTER II of INFRACTIONS AGAINST the GENETIC HERI...	2017-02-06 18:16:06
Copy Delete	98966	2895175	430f88843fd8c4930965f98528dc02b3	Art. 16. Access component of genetic heritage fo...	2017-02-06 18:16:06
Copy Delete	98967	2895175	924b4e25a50641213fee865279946714	Art. 17. Refer to the outer component sample of ...	2017-02-06 18:16:06

Fig. 3. Screenshot of one of the MySQL tables for penalties (Source: www.Global-Regulation.com)

After following the said steps, the system then note in its search instance whether or not a law has penalties attached to it, so that the search instance can filter by laws that have penalties (as opposed to Global-Regulation's regular search that includes laws that don't have explicit fines attached to them). This process is run as a batch job offline because 1.6 million laws takes several hours to process.

When a user does a search, the search is first sent to Global-Regulation's Elasticsearch instance, and then the penalties are looked up from the MySQL database afterwards. This allows full-text search of laws to be combined with penalties, and in a way that results in much less strain on Global-Regulation's relational database (because penalties are looked up by IDs rather than a JOIN). Storing the penalties separately allows to reduce the amount of data in the in-memory search instance, and decouples the services (since Global-Regulation have other types of search like technical standards and law analytics). See Figure 4 for the overall global penalties summary.

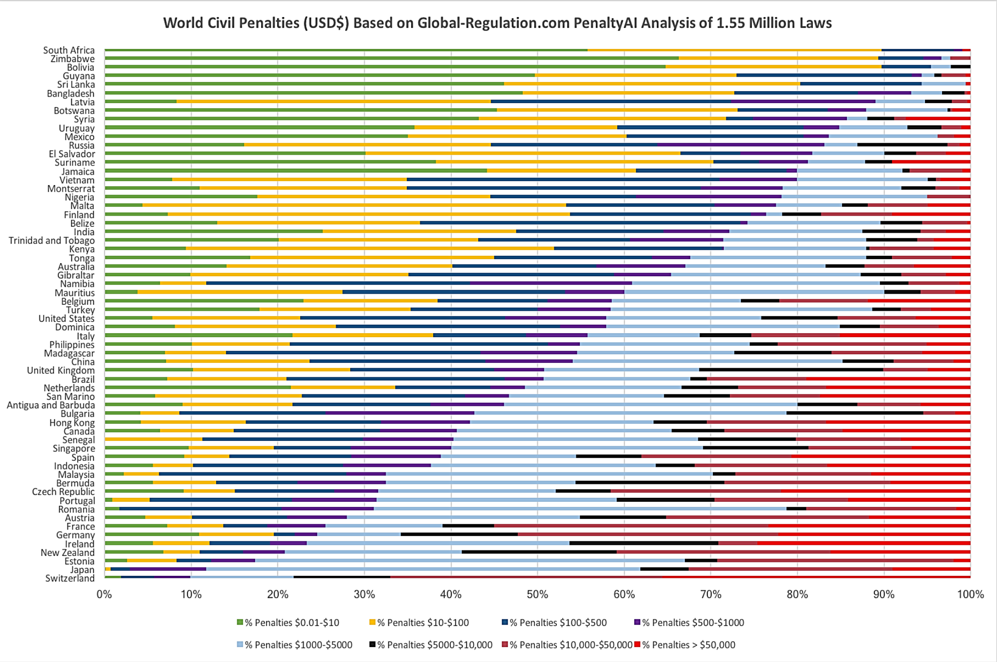


Fig. 4. Visualization of penalties for non-compliance (Source: www.Global-Regulation.com)

5 Concluding Remarks

Global-Regulation enhances linked democracy by using state of the art artificial intelligence technology to provide the world's legislation in English. Furthermore, Global-Regulation provides an advance system to extract compliance clauses and penalties from this legislation. Furthermore, Global-Regulation builds on the democratization of artificial intelligence and Microsoft's generous support to bring the benefits of technology in general and AI in specific, for the advantage of the public.

What could only be a dream few years ago, has turned with Global-Regulation's vision and advanced technology into an innovative tool of global democracy, now used by leading academic institutions and governments around the world. Going forward we look to expand this democratising tool in a way that will enable every citizen of the world to ask for her legal rights and responsibilities in her country of choice, based on this database of world laws and receive a reply, completely automated, with a click of a mouse.

We have learned that combining big data - the world laws, with advanced artificial intelligence, creates fertile ground for endless opportunities in the realm of bringing the law to the people and bridging the gap between society's bricks (laws) and its citizens. Yet the main challenge going ahead is twofold: how to use the technology in a way that will be understandable, intuitive and friendly to people; and, perhaps more importantly, how to explain to users around the world, that Global-Regulation is a key to linked democracy.

References

[1] Molly Schwartz, Democracy and open data: are the two linked?, May 22, 2014, Congressional Data Coalition, http://congressionaldata.org/democracy-and-open-data-are-the-two-linked/

[2] Dirk Helbing, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari, Andrej Zwitter, Will Democracy Survive Big Data and Artificial Intelligence?, Scientific American, February 25, 2017, https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/

[3] Democratizing AI: Satya Nadella on AI vision and societal impact at DLD, 17/01/2017, https://news.microsoft.com/europe/2017/01/17/democratizing-ai-satya-nadella-shares-vision-at-dld/#sm.00001d5vpu2vg5dm8tucb72g11vbl#iFfMr1YvJHXYZtg1.97

[4] Sean Goltz & A. Nikolic, Global-Regulation – Drawing Future Regulatory Tools from the Experience of the Past, 4(3) The European Journal of Risk Regulation 391-398 (2013), http://ejrr.lexxion.eu/list/articles/author/Nikolic,%20Aleksandar

[5] Alexis Lloyd, Disentangling Democracy From Geography, The Atlantic, May 9, 2017, https://www.theatlantic.com/technology/archive/2017/05/disentangling-democracy-from-geography/524124/

[6] Rebekah Monson, Freeing Technology From the Pace of Bureaucracy, The Atlantic, May 16, 2017, https://www.theatlantic.com/technology/archive/2017/05/freeing-technology-from-the-pace-of-bureaucracy/524034/

[7] Sean Goltz & M. Mayo, Enhancing Regulatory Compliance by Using Artificial Intelligence Text Mining to Identify Penalty Clauses in Legislation, MIREL 2017 - Workshop on `MIning and REasoning with Legal texts', held at the 16th International Conference on Artificial Intelligence and Law, King’s College, London, UK, June 12 - 16, 2017 [forthcoming].

Equal Access to Online Legal Information through Democratisation of Technology: A Myth?

Sue Ann Yap[1]

[1]JADE, Sydney, NSW 2000, Australia
Australia
yaps@barnet.com.au

Abstract. Wider access or "democratisation" of technology emerges as one of the most powerful force of change in the way we think, learn, exchange information and knowledge, and most importantly, in inspiring new ways to effect social change. In this paper, I will discuss that the advancements and emerging efficiencies in AI and technology has not democratised access to legal information. I argue that the innovation in technology has primarily been driven by private sector, to the detriment of the community at large.

Keywords: Innovation, emerging technology, legal publishing, equal access, open access to information, open access to technology, community-based outcomes.

1 Introduction

A normative, rational dialogue on democracy typically begins with Rousseau's definition that the state represents the will of the people through commonly agreed instruments and procedures. Terms such as "popularly elected" government, representatives, and "social contract" spring to mind (Rosseau 1767). In a simpler world then, if one posits that the developments in artificial intelligence and emerging technology serves to enhance exchange of information and ideas, break down geographical barriers and improve the human condition in general, then the reasonable (if not sanguine) assumption is then that a well-connected citizen can easily access and connect with information which enables the citizen to access public and legal information.

For this paper, I will discuss how access to the Internet does not necessarily mean equal access to justice, in the context of legal information; that the benefits from AI/technology innovation has been consistently usurped by market interests. I conclude with identifying that the cooperation and collaboration between private-public forces as the way forward in re-imagining equal access to the Internet *and* justice-related content.[1]

[1] I will be referring to artificial intelligence and technology inter-changeably. In this cluster, I am not differentiating between automation, process creation, big data mining, data analytics

2 Context and a *Reasonable Person's* Self-litigating Journey

Let me start this discussion with a context. In Australia, a person's most common pathway to justice is when law or regulation is being observed, as exemplified as applying for a passport or paying a speeding ticket, or even more mundane activities like getting married or registering a land title.[2]

I would like the reader to consider the following scenario:[3] you are enjoying a quiet picnic alone in the NSW bush land. Whilst enjoying your book, you hear a rustle ahead of you. You spot a fox pup, which seemed to be in distress. You sit and observe: not knowing what to do, for an excruciatingly long time. You tell yourself that you would wait for the pup's mother to arrive and you would leave when she does. That moment never arrives. The pup seems to be limping towards you and making horrible noises. You make the decision to take it to the nearest police station to seek help. You and your new companion, now warmly swaddled in your shirt in the passenger seat, are racing down the highway to the nearest police station. You are being guided by Google Maps, from the iPhone resting on your lap. A patrol car appears behind you, with sirens blaring.

You both stop at the shoulder of this highway. The police officer asks, "Do you know how fast you are going?" You nervously answer that you are speeding as you have an injured fox pup in your vehicle. By saying this, you have just admitted breaking the law. The police officer notices the iPhone on your lap and that you are shirtless, both of which are summary offences. He then proceeds to inform that you are knowingly transporting a game animal without a permit in your vehicle. [4] In a short span of time, you have just broken four different laws. The police officer, being the enforcer of black-letter law, issues you four tickets. One week later, you receive a Court Appearance Notice tor a CAN, to answer for your alleged "crimes". Until this moment, you have never broken any laws or even considered breaking any laws. The most mundane infraction you have ever committed is jaywalking in the city, which also happens to be a summary offence in NSW.

You decide to self-represent yourself in Court. After all, you did not knowingly broken any laws, you did not hurt anyone and no property was stolen. You are a victim of circumstances. You are confident that the officer of the Court would consider this matter with discretion. But to answer the charges in Court, you will have to form a plea or counter-claim against the summons from the State. Whilst you have not broken any aspects of private law, you have breached four separate public laws; laws,

and so forth. I will be using the over-arching concept of artificial intelligence and emerging technology as a development in the field of computer science, created by humans to effect efficiency, intelligent collaboration and waste reduction.

[2] S7 of the Australian Passport Act 2005 (Cth), Part 5.2 of the Road Transport Act 2013 (NSW), Real Property Act 1900 (NSW) and Marriage Act 1961 (Cth).

[3] Some of the material facts of the case are derived from a matter dismissed by the presiding district court judge in Sydney on 6 September 2016.

[4] Rule 300 of Australian Road Rules 1999 (Cth), Part 5.2 of the Road Transport Act 2013 (NSW), s5 of the Summary Offences Act 1988 (NSW) and s3 of the Rural Lands Protection Act 1998 (NSW).

which universally apply to every single person in NSW. As almost all government information is available online combined with a plethora of free material in the Internet, you are certain that you would be able to build your own argument. After all, you live in a free, democratic society with unfettered access to information.

As a non-legally trained individual, you invariably commence to locate court forms and procedures from the Court website and other non-profit sites which offer such templates. You perform searches[5] in Google to find out about your numerous offences. Thanks to Google's predictive text searching and pattern recognitions, you are able to locate what actual laws have been breached. You may even locate the legislative instruments and regulations, which detail the laws you have allegedly broken.[6] Your persistence and a stroke of luck may even direct you to types of remedies you are entitled to. Armed with this knowledge, you might like to find out if there are any cases in NSW or in wider Australia, which resemble your situation so as to assist in the building of your argument. Your searches, again operating with newspaper articles, opinion pieces, newsletters from law firms which specialize in traffic offences, and plenty more. There is a plethora of information: all of which seems to offer conflicting information. All that information, free and available: the dilemma is, which lot of information is credible and authorised for your day in court?

Suppose then, that this exact scenario is presented to a legal practitioner. This practitioner will review the legal issues and commence research to identify the rights of the clients and possible remedies. The research typically commences with Google, followed by any number of wikis, blogs, metasites and sites created by private and government entities. The practitioner, may even consult a specialist site and do research in a paid legal research platform like LexisNexis to locate commentaries.

Although we have the same scenario and juxtapose circumstance, the objective of this imagined journey is to understand this: how does an ordinary person access the legal information, which a person is entitled to in an open society? In simpler, terms, how does an ordinary person gain access to the justice system? And the question pertinent to this discussion is this: if technology is supposed to increase access to information, how is it the person in the first example above being disadvantaged?

3 The Internet and Accessing Legal Information Online: Themes

The examples illustrated in the previous section is not meant to trivialise the legal education and training that law students, and indeed legal practitioners who undertake years of studies to achieve their standing. Rather, it is a critique that the rule of law in its purest sense, does not seem to be served by the leaps in AI/ emerging technology. This observation is not unique to Australian. The fastest way in which any citizen would access "law" or the written word of law is to go online and search for infor-

[5] Free test searches which typically contain the queries or and not limited to "is it .." or "what is …" or "how is .."

[6] See www.legislation.nsw.gov.au

mation. In Australia, the state has been not only been providing and maintaining information for public consumption, but also by observing appropriate Creative Commons arrangement, thus enabling private citizens to reproduce and republish the information. To the best of its capacity, the Australian government (local, state and Federal levels) have attempted to meet its social contract with the community. However, the community stands to lose when the state is not pro-active and pragmatic in creating and maintaining tangible community outcomes when it comes to re-publication and reproduction of legal information: outcomes which include innovative consumption methods, value-added 'next-gen' apps and tools that would greatly enhance the experience in accessing legal information.

Azyndar *et al.* (2015:285) highlight that emerging online "next-gen" tools not only improves digital literacy, but it paves the way to the greater debate on how legal information can be ethically used and shared. They believe that the changing landscape of the legal publishing industry and indeed the legal profession in the US shows growing reception to smaller legal publishers: agile legal technology start-ups, adding value to public information and making value-added content available at affordable pricing models. Examples of such start-ups or 'disruptive legal technology' players in the US include Ravel Law,[7] Casetext and FastCase.[8] Azyndar *et al*. (2015) argue that the disruptors' entry into the legal publishing market pose significant changes in the legal profession itself: in the manner in which legal content is exposed to students to the evolution of new consumption and research patterns within the profession. More so, Azyndar et al. argue that the disruptors' involvement in the legal publishing market will pave the way towards a more equitable and affordable access model for the community at large.

The proposition that disruptive technology democratises the legal profession is also put forward by Yoon (2016). His article argues that emerging technology in access to legal information provides dual benefits: on one hand, it promotes greater access to legal information for ordinary citizens. On the other, enhancements in technology mean that lawyers can now discard the old economic model of practicing law, and begin to serve the rule of law in a more transparent, less routinized manner. In Yoon's

[7] Ravel Law was purchased by LexisNexis on 10 June 2017: a sale rumoured to be in the vicinity of U$20 million. See Venture backed Ravel Law sells to Lexis Nexis" in https://techcrunch.com/2017/06/10/venture-backed-ravel-law-sells-to-lexisnexis/, published 10 June 2017, accessed 12 July 2017

[8] Ravel Law (www.ravellaw.com) was founded by two law graduates from Stanford Law School in 2012, was one of the first, affordable legal content provider which models provides data analytics for the profession. CaseText (www.casetext.com) on the other hand, is case and legislation look-up service, founded by a litigator who found the legal publishing paradigm unaffordable and old-fashioned. CaseText received crowd funding from the legal community and raised U$7 million in 2015(see https://techcrunch.com/2015/02/03/legal-tech-startup-casetext-raises-7-million-series-a-round-led-by-union-square-ventures/). FastCase (www.fastcase.com) was founded in 2008 by a solicitor who found the cost to access legal materials prohibitive. All three legal techs, as they identify themselves offer free-to-air information available to members of public and an affordable pricing structure to encourage access to justice.

words, the legal profession returns to one, which serves the rule of law and the ideals of justice, and is no longer a luxury reserved for the wealthy (Yoon 2016: 66).

The self-litigating context that I laid out above, combined with Azyndar *et al.* and Yoon's propositions, reveal three themes that encapsulate the theme of this paper. Firstly, generally speaking, access to the Internet is affordable and has become increasingly easier. With Google as the first port of call, an ordinary citizen can pretty much conduct some form of legal research. There is no challenge or issue in entering the Internet in a liberal democratic country such as Australia (Cann 1989: 1168). Secondly, emerging technology and ease in which sites can be created means that whilst there is no shortage of legal information, forms, opinions, and blogs, there is no way to ascertain the currency and credibility of the sites. The Internet remains an unregulated sphere, powered by private interests and the adage that "In the web, no one knows you are a dog" rings true for the parties who consume, and the parties who produce or reproduce content (Christopherson 2007: 3038-56). And finally, which is corollary of the second point, that meaningful and useful information is often trapped behind an unaffordable pay wall. This highlights the dilemma: for all the easy and fast access to technology and information --- open access to legal information specifically, remains a myth. I will expand on these three points in the following sections.

4 Access to Internet and Unequal Access to Content

The normative argument reads that ubiquitous connectivity fosters faster exchange of ideas and easier access to information. And in turn, that the advancements as experienced in the Internet should lead to easier access to justice-related content. Various writers have argued that the reality remains that there is unequal access to the Internet and therefore, unequal access to content, meaningful content that is.

This view is argued by Cedar-Silva (2013: 17) when he posits that the Internet is a contradiction in praxis: on one hand, the advances and plurality of technology enables ease of entry to the Internet, on the other, the advancements have meant that private interests are gaining efficiency in finding newer, creative ways to monetise and even create higher barriers to access. This contradiction reflects the sentiment of Vin Cerf (2012) who maintains that the Internet is an enabler of rights, may those rights to be enforced by private or public interests.

Expanding Cedar-Silva's proposition, Lloyd (2001: 505) argues that unequal access to information or 'digital literacy divide' is at the core a socio-economic issue: that decreasing price of connectivity does not present equal opportunity to crucial social information, if the citizenry is still hampered by demographic, geographic and socio-economic factors. It is this socio-economic factor, which amplifies access inequality: that access to information and more specifically legal information is facilitated by the ability to pay. The implications for this passivity is chilling: that we as citizens and the elected government, are wittingly and willingly surrendering public information to market forces and allowing these forces to drive the innovation and modernization of public information. If the interests of the private sector drives innovation and content modernisation, then that means access to public information is reduced to a marketplace.

Cedar-Silva's contradiction is similarly observed by Perez (2013: 61-63), who argues that the success of the Internet is attributed to its democratic environment, powered by a combination of private-public interests intent to learn and create new ways of learning, production and consumption. Perez observes that as soon as an initiative is centrally coordinated or institutionalized, the creative process, consultation and civic focus is lost. Like Cedar-Silva, Perez argues that the Internet largely remains a private sector environment and the private interests will always be financially driven.

5 Innovation and the Commons

Innovation, especially in the field of AI/technology represents new opportunities for private sector to become more competitive and consumer focused, as well as enabling public sector to serve community needs in more efficient ways (Cann 1989: 1167). Innovation should mean the expansion of consumer choice and challenging the status quo of the commercial and public sector. However, if the rewards of innovation are not harmonized with the community at large, the drive for innovation will be to achieve financial gains. Cann (1989: 1168) argues that there is lack of recognition that innovation is predominantly driven by "parochial self-interest". He argues that it is flawed not to evaluate innovation in a more holistic manner: that is, harmonising the benefits of innovation for the community at large and the private sector. In the free market, states are reluctant to introduce protectionist policies to emerging technology (legal or otherwise) as it is seen to be paternalistic and stifling competition. It is in this climate that the private sector has the discipline and motivation to excel and bring forth the best of breed.

Add to the mix is the emerging trend of grass-root innovation. Growing interest and the low cost of entry to create online presence has witnessed innovative online tools created by private individuals. Specific to the Australian legal industry, the Internet is host to a plethora of independent websites. This demonstrates that the conversation has moved from "who has a right to internet" to "who has a right to legal content". To my point of changing consumption patterns, for the first time in a long time, New Zealand Law Reports, United Kingdom Reports, Victorian Reports and NSW Law Reports are all now available in pay-per-view format at affordable A$15 to A$25 per Case. These Reports are still available in an enterprise subscription format, but the availability of these Reports at this competitive rates indicate that the stakeholders in the Courts and Law Reporting Councils recognise the inevitably of change in the profession brought on by emerging technology and AI. In the last 3 years, the Australian and UK Law Report Councils demonstrated courage and commitment to positive social outcomes in creating and pioneering new consumption patterns apropos to the changing milieu.

Again, referring to the legal publishing industry, the private sector which used to and to an extent, still holds a monopoly on innovation, is now faced with growing competition from smaller, more agile independent players, a majority of them private individuals, who are able to create more agile ways of production, consumption and collaboration. The implication of this shift is simple yet powerful: the conversation about the benefits of innovation should now encompass the private sector, the private citizen and the government sector in a collaborative, reciprocal manner. The era of top-down monetization and consumption approach is rapidly losing its hold. The technological advances in connectivity mean that the dialogue about access to content is a tripartite conversation (Lloyd 2001: 505). The way in which we can translate innovation into improving social realities is one when the Internet is driven and advised by "participation of local stakeholders with a global reach in mind" (Cedar-Silva 2013: 26).

6 The Right to Internet as a Human Right?

The nature of the Internet is also its key success factor: that it is freehold, and that it crosses boundaries. It is this very characteristic, which makes the Internet, a prime post-modern means to improve social reality on a local and global level. No country on earth currently protects 'the right to the Internet'. No state or commercial entity can enforce his or her sovereign or private rights online, aside from the innocuous monetised enforcement. As such, the right to Internet is highly dependent on the level of development and democratic freedom in a specific country. This goes back to my earlier argument that whilst access to the Internet is fairly affordable, it is the equal access to content remains dubious. Article 19 of the International Covenant on Civil and Political Rights can possibly be used as a global covenant to safeguard this 'right to access'. More significantly for this discussion, Article 19 entrenches the freedom not only of access but also to

> "…to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice." (Land 2013: 393) (author's emphasis).

Although Article 19 does not guarantee a right to be online, it does provide a framework on how the government, private sector and private individuals can work to maintain this freedom. Land recommends the reading of Article 19 by recognising that the technology underlying the content, serves as "framing device" to enable choices to be made. That the Internet as an enabler of rights, is made possible through efficient architecture designing, easier to learn coding skills, increased open source standards and most importantly, a growing number of cross-border of grass-roots movements who collaborate to effect social change locally and globally (Land ibid, 395). I draw on the examples of the Arab spring revolution that was galvanized through Facebook and the resourcefulness of local activists in China who expose state censorship using other online mediums, as Facebook is banned (Cedar Silva 2013, 17).

How do the competing stakeholders in the online environment balance the public and private sector's interest to pursue optimum return on investment, in light of maintaining an entrenched freedom of access and freedom to access content in the format of choice? To this point, I ask the question: does that mean the traditional practice of placing key legal information behind an unaffordable pay wall, poses as a possible breach Article 19?

7 Public and Private Interest serving to Democratise Access to Technology and Access to Content

In light of the dire diagnosis of innovation and benefits of technology, is the common person doomed to be the last adopters of due to high barrier to entry? As cheaper access to Internet feeds the growth in self-taught programmers and grass-roots' sites, it is realistic to utilise a neocorporatist model in engaging the three parties in the dialogue on democratising access to technology and access to content (Barton 2015: 542). Without diverging from our discussion, it is important to remind ourselves that the neocorporatist or liberal corporatist approach describes how the socio-economic policies of a society is brought about through the engagement of private sector, private individuals and the government through consultation, cooperation and establishment of common benchmarks, protocols and outcomes (Preminger 2017: 85-99). This may seem idealistic, but this approach will mean that every stakeholder level not only has a say but has an obligation to effect a change in the commons: may it be in enforcing civic rights or pursuing commercial pragmatism. If the public sector and private citizens remain a passive voice in 'netizenship' then the paradigm will always be one, which prioritises commercial interest. (ibid: 90).

I do not suggest that in order to reclaim the Internet as a civic commons, one need to embark on radicalized movements as exemplified by anti-globalised movements. Rather, it is a dialogue towards establishing tripartite protocols involving the nomenclature from all sides to enable improved access to the Internet and access to content (Waters 2004: 854-874). The following are my proposed re-imagining on democratizing access and access to content, in the context of legal information.

7.1 Sustainability and maintaining the quality of independent sites:

Returning to the 'fox pup' scenario and access to legal information in Australia, content is predominantly found in government sites as well as commercial online sites maintained by LexisNexis, Thomson Reuters and Wolters Kluwer. For the members of public, legal content can be sourced from the publicly funded, non-profit met site AustLII, JADE a crowd-sourced, freemium legal site and a plethora of other smaller self-publishing non-profit sites. Irrespective of the public-private nature of the sites, all raw data is sourced free-of-charge from the state.

The breakthrough in publishing technologies and e-business innovation as produced by the commercial publishers are hidden behind paywalls. It is ironic that the innovation based on data sourced from the state; yet the state does not benefit from this data exchange. In fact, consider the millions of dollars which have to be spent on legal publishers by the State Department of Justice annually to subscribe to reports, case law and legislation.

In the absence of any publishing protocol or regulation, the ‘revolution’ of the smaller, online publishers take form: some for altruistic reasons, some for commercial reasons.[9] The unaffordability of the commercial legal publishers and low cost to data parsing has witnessed a demand (and support) for smaller publishers. In Australia, an increasing number of practitioners (and non practitioners) are turning to AustLII and JADE. These two online publishers, one a donation-based metasite and another crowd-sourced funded site, are constantly monitored by Judgment Offices for accuracy, currency and quality. This leaves a gap in maintaining such standards in the commons. In the absence of a coherent policy or republishing protocol, the relevance of sites who self-publish are defined by their own purpose. A donation-based site is behoven to foundational funds, a crowd-sourced organisation can only grow as fast as the crowd-funding metrics and a privately funded site will only last as long as the entrepreneurial drive is sustained (Cay Johnston 2006: 65) There is an imperative to create a set of republishing standards which maintains accuracy, quality, currency, access and technology knowledge exchange.

7.2 Community-based consultation, outcomes and benchmarks

The *raison d’être* of the private sector is to monetize and increase shareholder value. This is predominantly done through identifying innovative ways to produce and consume. The public sector does not hold this reason: its primary reason of being is to serve the community. The public sector is not expected to be agile. So, in the context of legal publishing, how can the community at large benefit from any publishing agility or innovation? As there is no reciprocal technology transfer protocol between private sector and the government, the state is literally giving away the family jewels to the private sector in the absence of community consultation. I propose that in this context, that the community is informed on the manner in which legal information is harvested from the government sites, the manner in which they are republished and commercialised, and the effected ‘returns’ to the community. Private citizens have a right to know how their taxes are being used to inadvertently fund private innovation, and have a say on how this can benefit the community.

Further to this, the public sector can no longer rely on the private sector to realize change but it must actively engage with smaller entities to create a monetization, commercialization outcomes and reproduction policy, which provides equal playing field for all.

[9] In 2012, US legal tech start-ups or disruptive legal technology providers raised over U$60 million and 2013 witnessed a growth of $458mil. This figure started to decline as cost of programming and cloud computing become more affordable, as reported by Joshua Kubicki, 2013 Was a Big Year for Legal Startups; 2014 Could Be Bigger in Tech Cocktail (Feb. 14, 2014, 12:07 PM), http://tech.co/2013-big-year-legal-startups-2014-bigger- 2014-02,

An example of successful and profitable private-public sector collaboration is the arrangement between CanLII, metasite similar to AustLII based in Canada. CanLII's content is powered by Lexum, legal technology provider based in Montreal which created a 'freemium' model to law reports in Canada. This site is not funded by donations, rather through a combination of funds raised by law societies and the members of the Canadian bar. Lexum similarly collaborates with other disruptive legal technology players and from an overseas standpoint, seems to co-fund the metasite through innovation-transfer.[10]

7.3 Platform agnostic and blockchain transactions

Blockchain, in the simplest terms, is a system of recording digital transactions: from financial records to how many times a PDF was downloaded from a monetised information website.[11] The underlying structure of the blockchain reveals a network of distributed databases capable of reconciling transactions. Consider the legal publishing world: a new paradigm in which the user does not have to subscribe to any platform but is still able to purchase and access content from across a myriad of platforms. The user does not have to consider currency or inventory exchange or how bills are issued. The user simply searches and purchases content at the end, and the transaction is reconciled. The user receives a bill on the frequency required i.e. monthly or quarterly, and the publishers of the content are monetised as soon as the purchase takes place.[12]

The key to blockchain resembles the milieu that I referred to earlier: that it operates in publishing protocols and standards agreed to by the private-public and community. The hard question remains: despite the growing trend towards platform agnosticism and Google-pendency, no commercial legal publisher of content would allow federated searches to cross their pay walls. All publishers, legal and commercial believe in the superiority of their content and the expensive, unaffordable and bewildering pricing structures reflect this perceived superiority. The lack of collaboration between publishers also reflects their fear of losing market hegemony and the disappearance of smaller disruptors gobbled up by capital (Gallacher 2009: 8-10). Thus, returning us to the debate of unequal access to content despite equal access to the Internet.

[10] https://lexum.com/en/ accessed 19 July 2017

[11] Definition of blockchain from https://blockgeeks.com/guides/what-is-blockchain-technology/ and "What is Blockchain" from http://www.coindesk.com/information/what-is-blockchain-technology/, accessed 19 July 2017

[12] At the time of writing, the Consolidated Councils of Law Report in Australia, LexisNexis, ThomsonReuters and JADE have agreed to an interlinking protocol which enables a user to link from one platform to another. This groundbreaking protocol represents the judicial and practitioner recognition of enabling equal access to content, and is the first concerted step in private-public collaboration. Aside from JADE, none of the legal publishers have opened this option to their customer base. AustLII and Wolters Kluwer have declined to participate in the undertaking of this protocol.

8 Conclusion

The innovation and development in AI/technology and indeed the cheaper access to the online environment has not provided ‘netizens’ with equitable access to legal information. The ‘netizenry’ remains a disgruntled yet passive voice, happy to grouse that the state is failing its responsibility in sustaining and maintaining equal opportunity to technology and content, but also unwilling to be involved in effecting change. The paradoxical paradigm of cheaper access to the Internet and equitable access to legal information can only be addressed if the three major parties in this environment recognise the importance of each other’s role and obligation in fostering innovation and equitable access. The private citizens, by being silent and passive on the commodification of public information is as culpable as the state in eroding basic human rights. The private sector, unimpeded by government policy or community-based protocols will continue to pursue monetization objectives. To close, consider a quote from Lloyd (2001) who said, “We must understand that the digital divide, like the justice divide, is a political divide”.

References

1. Azyndar, S., Lee, K. and Mattson, I,: A New Era: Integrating Today’s ‘Next Gen’ Research Tools Ravel and Casetext in the Law School Classroom. Public Law and Legal Theory Working Paper Series, 285 (2015).
2. Benjamin H.Barton: Some early thoughts on liability standards for online providers of legal services. The Hofstra Law Review, Vol 44, pp. 542 (2015).
3. Bertoni, E.: Towards an internet free from censorship: a proposal for Latin America, University of Palermo (2012), paper authored for Center for Studies on Freedom of Expression and Access to Information, Universidad de Palermo. Sourced from https://citizenlab.ca/2012/03/towards-an-internet-free-of-censorship-in-latin-america/ on 9 July 2017.
4. Cann Jnr, W.A.: The depoliticization of takeover theory: creation of an innovation factor. Syracuse Law Review, Vol. 40, pp. 1167-1168 (1989).
5. David Cay Johnston: Giving Charities a Voice: The Legacy of Norton J. Kiritz, The Chronicle of Philanthropy, Apr. 6, pp. 65 (2006).
6. Cedar-Silva, A.B. : Internet freedom is not enough: towards an internet based on human rights. South African Journal of Human Rights, Vol 18, pp. 17 (2013).
7. Cerf, V.G.: Internet Access Is Not a Human Right, N.Y. TIMES, Jan. 5 (2012).
8. Christopherson, K.M.: The positive and negative implications of anonymity in Internet social interactions: On the Internet, Nobody Knows You’re a Dog. Computers in human behaviour, Vol. 23(6), pp. 3038-3056 (2007).
9. Gallacher, I.: "Aux Armes, Citoyens!"Time For Law Schools To Lead The Movement For Free And Open Access To The Law. Toledo Law Review, Vol. 40: 1, pp. 8-10 (2009).
10. Land, M.: Toward an International Law of the Internet. Harvard International Law Journal 54, 3, pp. 393 -395 (2013).
11. Lloyd M.: The Digital Divide and Equal Access to Justice. Hastings Communications and Entertainment Law Journal, Vol. 24, pp. 505 (2001).

12. Perez, O.: Open Government, Technological Innovation, and the Politics of Democratic Disillusionment: (E-)Democracy from Socrates to Obama Journal for law and policy for the information society, Vol. 91, pp. 61-63 (2013).
13. Jonathan Preminge: Effective citizenship in the cracks of neocorporatism. Citizenship Studies, Vol.21:1, pp. 85-99 (2017).
14. Rousseau, J. J. The Social Contract, Translated by Christopher Betts, in a volume entitled: Discourse on Political Economy and The Social Contract, Oxford University Press (1994)
15. Waters, S.: Mobilising against Globalisation: Attac and the French Intellectuals, West European Politics, Vol. 27: 5, pp. 854-874 (2004).
16. Yoon, A.H.: The Post-Modern Lawyer: Technology And The Democratization Of Legal Representation. University of Toronto Law Journal, 66 (2016).