INTERNATIONAL COMPUTATIONAL LAW JOURNAL

CODEX SOURCE ARCHIVE

Democratizing the Law with Open Data

Gabriella Capone

Computational Law Report · Release 2.0 · 2020-11-20

Reader edition prepared by The Research University under CC BY 4.0. This cover was added; the original PDF pages follow unchanged. Original authorship, publication identity and source links are preserved. This independent republication is not sponsored or endorsed by Stanford University, Stanford Law School, CodeX or MIT.

Volume 1 · Issue 2 · Systems, Rules and Public Infrastructure

Original source and license evidence are recorded on the article landing page.

MIT Computational Law Report

Democratizing the Law with Open Data

Gabriella Capone

Published on: Nov 20, 2020

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

ABSTRACT

Open data is not only integral to computational law applications that harness accurate, public, legal information, but it can also underpin better justice and legal outcomes. Open data is a key to accelerating the law itself. Exciting efforts show legal data impacting legal outcomes, judicial efficiency, access to justice, and more. Often, such efforts often utilize data provided via ambitious private and academic efforts, for rich, public portals of legal information remain few and far between. This piece explores opening justice systems, with a focus opening data from federal and state legal institutions. It assesses the state of play in these systems enabling via legal frameworks, institutional capacity, and already-opened data. In particular, it calls for further study of state justice systems, in which the majority of legal actions in the US take place. Yet, gaps in these systems also represent opportunities. To support computational and equity-driven efforts, the piece discusses barriers to opening legal data, and potential breakthroughs. Beyond the benefits of Open Government, open legal data is foundational to a dynamic legal system.

0. Introduction: Why open legal data and why now?

In the same way that law serves as the foundational input by which our legal system has traditionally operated, data serves as the essential input for understanding, mechanizing, and analyzing the law in an increasingly automated fashion. When available at scale, legal data can benefit legal outcomes and justice systems immensely. Academics can use administrative and case outcome data to uncover systematic bias in rulings to promote fairness. Practitioners can analyze precedent, sentiment, and judicial decisions to identify patterns common to winning cases, with the potential to democratize best practices. Citizen scientists can assess efficiency and effectiveness of legal counsel by comparing courts caseloads and outcomes, towards improving resource allocation. Additionally, advocates can test interventions to improve criminal justice outcomes, providing invaluable insights for policymaking. Such examples only begin to scratch the surface of what is possible with legal data.

At the systems level, data is a key mechanism for innovating the law. There is room for opportunity in this respect. The standard legal curriculum looks the same as it did in the 1800s. There is no central, digital repository of the laws in the US, with the private sector-academic collaboration Case.Law being the closest contender. More often than not, attaining legal documents requires appearing in-person at the court in question, and paying to print copies of the documents. If a federal case, you may pay a fee-per-page download (employees of the US Department of Justice are no exception to this rule). For those courts and departments of justice that are online, practices and formats are divergent; the key policy for bringing state legal information online does not prescribe uniform technology practices. Rarely can you do things like access an API for court docket information or analyze a jurist's caseload file to study outcomes.

These points start to illustrate a picture of the current landscape for law and data: it is hard to bring empirical and data-informed practices into the law. This gap is particularly acute if you seek primary, public, authoritative legal information, such as: case-related documents, dockets, opinions, outcomes, caseload statistics, administrative information, case law, and more. As we will see, tech-enabled efforts are underpinned by largely private and academic labors to generate data central to the mandate of — but not provided by — public institutions. For perspective, observers estimate that less than one-third of US datasets that can be made available to the public are fully open.[1]

While open legal data can directly enhance system efficacy, efficiency, and equity for something considered by many to be a foundation of government, there are several notable second order benefits that underscore the value of this practice. Open legal data can also support applications in computation, data science, and empirical study at scale. Ample amounts of high-quality data enables accurate representations of referents — processes, individuals, systems, bodies of law, or otherwise — to build high-fidelty models or comprehensive, automated processes. Data from public legal institutions can also provide requisite information for human-centered law and design to characterize service delivery, fairness and equality before the law, and access to justice. Further, in the spirit of impact- and goal-oriented approaches, data is necessary to track and measure outcomes, towards understanding and iterating system design. Finally, coordinated opening of legal data can facilitate normalization and interoperability of systems to bridge designs — unifying idiosyncratic methods, bodies of law, languages, formats, and more. Fundamentally, citizens ought to have access to the laws they are to follow; in 2020 this should mean true open access.

Indeed, one of the most promising developments to come out of the COVID-19 pandemic is a glimmer of hope for the digitization and modernization of legal institutions. The Supreme Court is hearing arguments remotely and live streaming oral arguments for the first time in its history. Michigan now offer "virtual courtrooms," enabling continuity of existing proceedings. Virtual arbitration and hearings are also picking up steam. That the Supreme Court held earlier this year that the official code of the State of Georgia is "ineligible for copyright protection" is another exciting boon to opening up legal institutions, beyond virtualization. It is unclear whether these changes will stick, but they do show adaptability towards coming online.

At the same time, protests against police brutality and systemic racism against BIPOC Americans highlights the value of open law and justice information. Waves of work examine the scope of law enforcement and administration of justice as it affects different parts of the citizenry. Citizens are galvanized to understand the administration of justice and how to improve it, especially for our most wronged and vulnerable populations. If we remain vigilant and work together to develop a culture where data is shared, these episodes can coincide to open the law more than ever before — not by

circumstance, but due to a deep-seated need. Information from public institutions is integral to solving pandemic-level problems and to meeting the justice challenges faced in America and around the world.

This piece looks to understand, promote, and advance the openness of legal institutions. Within the spheres of open government and open justice, open data can enable computational and forward-thinking legal efforts. In parallel, open data can enhance advocacy, by providing the basis for achieving greater equity and justice. While a great deal of legal information is already available to the public, it is often not "open" in the true sense, signaling issues of institutional capacity: the lack of a mandate, expertise, resources, or impetus to open game-changing information to communities. The rest of this article proceeds to survey open legal data in the United States as follows: first, by looking at judiciary and departments of justice at the federal and state levels. Then, the analysis examines compelling use cases for the data in the open justice and related domains in order to demonstrate the some of the broader relevance and impact of open legal data. Finally, this article concludes by examining some of the barriers and breakthroughs that are needed in order to open more legal data.

1. The 50 States of Data: A survey of federal and state institutions

An exciting trove of potential open legal data[2] resides within public legal institutions. This survey will examine the current landscape in order to better understand the gaps, opportunities, and design improvements that could benefit federal courts, the US Department of Justice (DoJ), and state courts. Specifically, this section will explore the following:

1. **Key laws and policies** that serve as the framework by which open legal data can be generated. Such mandates usually include a forcing mechanism and may even provide resourcing as a means to boost capacity.
2. **Institutional capacity** may include talent with expertise in data management: from normalizing, digitizing, structuring, provisioning, and maintaining information. Technological capacity is also critical; provisioning, hosting, and maintaining information online requires infrastructure. Sufficient funding is also a key piece of the puzzle here.
3. The **current state of openness** assesses what is available per the Open Definition, with a focus on: machine readability, accessibility online, and freedom for use by anyone for any purpose.

a. Federal open legal data: DoJ and US Courts

A strong network of laws and policies, as well as technical infrastructure, puts the DoJ in a strong position to open information. However, its opening is happening at a relatively slow pace. Compared to state and local courts, US Courts are overall less open, given a weaker mandate and higher incidence of opaque practices.

Key laws and policies. Open data builds on freedom of information and copyright laws. Familiar to open data advocates, the Freedom of Information Act (FOIA) of 1966 opened the door to public requests of qualifying information from federal, executive agencies. FOIA is a key avenue for accessing legal information: DoJ received nearly 100,000 requests in 2019, with two of the top five most popular reports on FOIA.Gov pertaining to DoJ data.The FOIA Improvement Act of 2016, subsequently passed into law, operates under "a presumption that such records [made available under FOIA] should be accessible to the American public," and provides that information be made available "in an electronic format." The 2016 amendment bolsters the digital provision of DoJ data through its format mandate.

The Obama administration propelled a number of advances towards opening federal agency data, including the Digital Accountability and Transparency Act of 2014 (to open federal financial data), the Open Government initiative,[3] and related projects, such as Data.Gov and Project Open Data. Strengthening both FOIA and Open Government,[4] such policies helped spark a movement, including the founding of the Open Government Partnership, with over 75 country-members. Then, in 2019, Congress passed the Open, Permanent, Electronic, and Necessary (OPEN) Government Data Act, codifying Obama-era work. The OPEN Act applies to federal agencies and institutes a policy of machine-readable and open by default for government data.

The Judiciary is effectively untouched by the Obama-era policies. Per US copyright law, works published by the US federal government cannot be copyrighted and must be placed in the public domain.[5] Such information cannot be copyrighted or charged at a cost — unfortunately a familiar barrier facing data-seekers at state levels. Interestingly, this copyright law also provides the basis for placing legislative and judicial content in the public domain. However, there is no requirement that information be digitized or open.

Institutional capacity and openness. Given an established set of platforms of accessing federal government data, the following looks at capacity and openness for three key platforms for accessing information from US DoJ (Data.Gov) and US Courts (govinfo.Gov, PACER).

i. Data.Gov

The central repository for federal data, Data.Gov also covers select municipal, state and tribal government information in over 200,000 opened datasets. It is a key technical resource available for federal agencies to open data. While Data.Gov was trail-blazing when launched in 2009, in the past decade the quality and quantity of open data on the platform has lagged. First, there are few datasets being added to the program. In 2017 there were 200,000 datasets and in September 2020 there were 207,000 — modest growth with the caveat that earlier in June 2020 there were 210,000 datasets (3,000 removed over the period). While existing datasets are updated, annual statistics from the past three

years confirm a reduced rate of opening up data. There is also a chance to open data from 40% of the qualifying agencies that have yet to upload a single dataset.[6]

While DoJ has met these bars and Data.Gov is a satisfactory repository for criminal justice data, the quantity of data provided is relatively low, referencing a narrow subset of topics. Of 150,000 federal datasets provided by 29 agencies, 1,200 come from DoJ, of which 64% are public — just under 800 open datasets. It primarily provides summary and census criminal justice data, including annual parole surveys and census of state and federal correctional facilities, recidivism, arrests and parole studies. The DoJ has also opened various types of information, including annual reports, Antitrust Division case filings, Americans with Disabilities Act reports, and administrative reports. Datasets are updated in varying degrees over time.

While the data provided does meet openness standards, the availability is waning over time and, due to broken links and reference to third-party sites, compromises accessibility. From a capacity perspective, a number of datasets — over 80% — are hosted by an external institution, raising the question as to expertise, resourcing, and technology to open data over time.[7]

ii. govinfo.Gov

Run by the Government Publishing Office, govinfo hosts official federal law and provides access to official publications from all three branches. In terms of judicial information, the Administrative Office of the United States Courts publishes, in text searchable format, written opinions issued after April 2005 from 205 district, bankruptcy, and appellate courts. Govinfo even lets users search by the nature of the suit. With this, govinfo fully meets open criteria and is a strong example of opening and maintaining legal information (in contrast with Data.Gov). Search capabilities and transparent management strengthen the offering. By fall of 2020, the site handled 188 million retrievals and added tens of thousands of new documents annually, at an accelerating rate.

iii. PACER

Operated by the Administrative Office of the Courts (AOC), PACER houses over one billion federal court documents: case and docket information from federal appellate, district, and bankruptcy courts, as well as administrative and referential data,[8] available as .html or .pdf. Records date to 1999, with older searches requiring a call to the court where the case was filed.[9]

PACER is generally not free to use. Written opinions are published on the relevant court website and free on PACER. Beyond that, all other case information costs $0.10 per page, up to $3.00 per document, matching the printing charges when visiting the relevant court directly. This presents a significant barrier to access for civil society, journalists, and even for federal government agencies looking to access information. The Free Law Project found that between 2010 and 2017, the DoJ paid nearly $25

million in PACER fees. In 2015, this amounted to $145 million in revenues, and over one billion in revenue over the preceding 20-year period. The AOC states that the fees are charged to cover platform maintenance (the US Courts cannot profit off PACER), which provides the resourcing to host and maintain the platform, but at the cost of openness. Pushing PACER to the "closed" end of the spectrum are policies that discourage users from gathering data: users who create "an unacceptable level of congestion or disruption to the operations" may risk account suspension. This may, again, be related to cost constraints around hosting but nonetheless violates open principles. Fortunately, the Open Courts Act of 2020 targets judicial record-keeping, and seeks to consolidate federal records into a single system and make them freely accessible to the public.

b. State-level open legal data: state courts

The majority of civil cases in the US are brought in state courts: roughly 95%. Over 90% of those incarcerated in the US at a given moment were convicted for violating state criminal law. While there is a legal basis across all states to make information publicly available, capacity to bring legal information online varies broadly. Ultimately, the openness of this data is highly variable across jurisdictions, and weak in aggregate.

Key laws and policies. There is no mandate for states to put legal information into the public domain, or even to digitize it. However, every state has some form of a freedom of information law to promote availability of state information, like sunshine acts, right to know laws, open records laws, and open meetings acts. Over half of states have *some* form of open government regulation. With that said, such information can vary in openness. For example, non-state citizens may be prohibited from accessing information,[10] or access may be limited only to lawyers who are members of the bar in that state. However, at least ten states do have open data policies that emanate through some combination of legislation, executive order, or both. Unfortunately, in recent times, the push towards passing such laws at the state-level has waned.

For states that opt-in, the Uniform Electronic Legal Material Act (UELMA) provides structure towards opening legal data, even if the open standard is not perfectly met. Adopted by 20 states, the UELMA seeks to give published online legal materials, "the same level of trustworthiness traditionally provided by publication in a law book." With the issue being that legal information consulted online may not be authoritative, or even accurate. Coverage via the UELMA is to include: the state constitution, state session laws (acts), codified laws (statutes), and agency regulations that have the effect of law. States can go further and include additional legal materials, such as the administrative register or judicial opinions. Official, electronic, legal material published by adopting states must be authenticated, preserved, and accessible for use by the public. This largely matches the open standard although machine-readability is not provisioned. Nor does the UELMA prescribe a particular technology for states to adopt, which could promote standardization and interoperability. Though,

ambitious states may emulate UELMA-adopter D.C. hosting legal code on GitHub (advocates can also help in the undertaking).

Institutional capacity. In terms of technical infrastructure, every state has some form of an online data portal. Topic coverage on these platforms is not necessarily as broad as is Data.Gov. For instance, Open Georgia only provides budgetary information, mostly via .pdf, whereas Colorado provides a broad topic coverage with nearly 2,000 datasets in varied formats. Nor should users always expect a large quantity of datasets. Pennsylvania makes just under 400 datasets available, while New York City alone provides 2,200 datasets. Wyoming's principal open data portal focuses opening geospatial data. Institution-specific portals can be more specialized, but state portals can aggregate, normalize, and unify data for access. Data.PA.Gov has a handful of datasets from the Administrative Office of the Pennsylvania Courts, showing how state portals can reduce a practical, technical barrier to opening.

To assess the capacity and openness of state justice systems I use a 2016-2017 survey of 232 courts from the National Center for State Courts (NCSC). Given the variability across state systems and barriers granular research, the review is done in aggregate.[11] The survey is complemented by legal scholar Sarah Glassmeyer's State Legal Information Census: a comprehensive, 14-point scoring of the openness of state legal information.

Of those courts reporting on remote public access capabilities, over 75% can provide remote public access to case-related information, including scheduling information and case documents. This existing technological capacity bodes well for opening data although it is not clear how the information is provided. For instance, Arkansas hosts court-related data via CourtConnect portal via the Contexte Case Management System (CMS). Even still, hosting schemes vary and may require extra-institutional capacity. The Colorado Judicial Branch does not host any data, but refers users to the court where a case was filed or to third-party vendors, such as LexisNexis, to which users can pay $7 per search.

Openness. When it comes to case-related information, 37-54% of respondents already make the data available for free online, depending on the nature of the data. This is consistent with Glassmeyer's finding: "For the most part, accessing state published law via the Internet is a cost-free experience for the user." Some 21 states provide some form of remote, online, public access to court records. At the same time, Glassmeyer finds: "no state provided barrier-free access to their legal information," and that across most states "it is impossible to do any but the most basic of legal research for free using state provided legal information sources." Of 152 courts providing public access to case information, about 25% require users to pay a fee depending on the information, namely because the systems are operated by contractors.

When it comes to accessing legal data, only one state of 21 surveyed allowed downloading of publicly-available electronic files without restriction, with hesitation of files released en mass being a chief

concern. Ambitious non-public efforts have started tackling this challenge, including the Caselaw Access Project and Measures for Justice, both of which are building a picture of state-level criminal justice data for county-level comparison and analysis.

While there is a long way to go in getting the sort of access that is deserved, there is reason for optimism, even from this limited sample. First, a majority of courts have the requisite infrastructure to host open legal data; in cases where this information is not free, there is a question of removing the fee to make it more accessible. The minority host information online but do not make it available to the public (5-23% depending on type); this demonstrates a clear way forward: removing access fees would be a win for openness.

Departments of Justice are also an important source of legal data to unlock. For instance, California's Open Justice, provides a range of criminal justice datasets that also promotes community exploration via a forum, and spotlights exciting analyses that have used the data. The data is open: free, machine-readable, available for bulk download, and completely open for use. There has been little study of state-level DoJ openness, though Open Justice is a promising example of what is possible.

2. Inspiration: five applications to demonstrate the potential of open legal data

The next section analysis focuses on the potential of what open legal data can enable. It is the position of this article that open legal data can enhance legal and justice systems by enabling novel applications *in* law and *beyond* law.[12] Because it promises to mechanize, automate and accelerate legal applications, one area, in particular, that stands to benefit from open legal data is computational law.

a. Support computational applications

Data from and about the institutions that make up the legal system is invaluable to computational applications. Previously in this publication, Professor Sandy Pentland outlined key components to enable computational law systems: goal definition, metrics and evaluation, testing, adaptive systems design, and continuous auditing. Perhaps fundamental, but legal data can help computational law to flourish by providing rich, authoritative, accessible information at scale to plug into any analysis that touches upon legal institutions. As the rest of society continues to develop new and innovative solutions, the need for open legal data will only continue to increase.

Ravel Law, a US company, is evidence of the power of data about public legal institutions. Ravel is built on the Harvard Law Libraries Case.Law platform, which it helped populate with over six million pages of legal texts (thanks to Harvard's extensive collection rather than accessibility via courts themselves). Ravel enables customers to perform new types of analysis on key aspects of the court system that would not be possible without court record being represented as data. Relying on

centuries of case law, users can compare Forums and outcomes or identify key precedents, terms, and judges influential in the case context. In practice, this helps lawyers better represent clients, from forum shopping to building stronger arguments. At scale, these types of insights can accelerate effective legal practice.

Administrative data and metadata are full of potential, too. By drawing on novel approaches from other domains, metadata analysis on legal datasets may lead to meaningful interventions. For instance, openPDS is a "framework that allows individuals to collect, store, and give fine-grained access to their metadata to third parties" analyzed mobile phone metadata (e.g., location data, phone call logs, and web searches) of veterans in aggregate in order to identify patterns of healthy and unhealthy behavior so that interventions can be made for the veterans more susceptible to experiencing PTSD.

b. Identify legal problems to shape legal services

Open legal data can shed light on legal problems, helping to identify legal needs, including: ability to pay, housing, violence and crimes, family disputes, and employment and business, enhancing the delivery of support to underserved groups. The Court Statistics Project analyzes caseload data to quantify self-represented litigants, shedding light on underrepresentation. Recidivism prediction is an increasingly active field. For instance, organizations and states can design better criminal justice systems by using state data to improve interventions. Recidiviz provides states and nonprofits with tools to measure impact of corrections and reentry programs, compare outcomes, and safely reduce incarceration in the US.

In turn, public institutions can better deliver legal aid, and shape processes and policies. Scholars at Stanford Law School highlight exciting access to justice efforts that harness legal data. Learned Hands gamifies the identification of legal problems, towards building a stronger model to detect legal issues that can be used for broader applications. Code for America's Clear My Record works with California counties to expunge criminal records so that legal defendants and less-resourced legal services may be able to empirically determine the best forums in which to bring claim types, the history and inclinations of judges, and the most significant precedents.

c. Improve engagement with the legal system

Open legal data enhances advocacy by opening raw information, and by reducing the expertise and specialization barriers.

Through Community.Lawyer, individuals can build modular, no-code legal apps that help automate legal processes — from citizenship applications to filing a fee waiver in a debt collection suit. This application shows how lawyers and technologists can make legal knowledge more accessible to communities. With accurate information funneled into such platforms, it becomes easier to navigate

local, legal situations. Further, by creating an open marketplace where different legal applications can be shared and iterated, best practices can begin to emerge and, as a result, it is possible to generate powerful network effects that advance the legal culture forward.

These new and dynamic ways of interfacing with the legal system also mean that it is easier for non-lawyers bring new perspectives to the legal system by engaging with legal data rather than needing to broach more complex or inscrutable legal processes. Broader access can raise important justice-oriented questions from communities and support conclusions. Beyond effective interventions, in line with Recidiviz, advocates can characterize and address bias in judicial decision-making towards improving fairness, using court data to assess patterns around jurist and defendant race, ideology, and more. In fact, open legal data directly aligns with social movement lawyering that contemplates how the law can inform advocacy efforts: it is not just statistical or textual data, but a way to capture the entire system for broader study and change.

Crowdsourcing science affirms that innovative discoveries do not require participants to be an expert (or, in this case, to have a law degree).[13] One study shows that a team of individuals of "average ability" could outperform expert teams. These individuals brought local expertise to bear on the problems, and generated more innovative solutions than did experts. Coupled with proper structure, the general public contributions can exceed those of experts. Successful crowdsourcing at NASA affirms that highly technical, public organizations can benefit from more open practices. Following a 45% budget cut in 2005, NASA had a dire need to "expand existing search capabilities for novel technologies" for space flight. Inspired by the TopCoder Challenge, it set out seven challenges to generate novel technical solutions.

d. Improve the efficiency, efficacy, and equity of the legal system

Legal institutions stand to benefit from openness, just as other institutions have. When a D.C. resident did not receive a summons as part of eviction proceedings started by his landlord, he was compelled to understand the system and if it might be impacting other tenants. He discovered that tenants are to receive summonses by private process servers (hired directly by landlords), and that, at best, such servers reached 60% of tenants. For one server, out of 1,660 cases, only 11 tenants were served. Additional defects were uncovered and stemmed from particular servers, and the overall effort highlights the misaligned incentives of notification and the extreme challenge facing evicted residents: an automatic forfeiture if they are not served and made aware of proceedings filed against them. By systematically sampling the delivery of summonses, journalists revealed a gap in the reporting and results of process servers. This type of revelation can directly inform policy, improving process design and citizen experience. Not to mention what such studies could do to improve organizational and operational processes.

Shifting judicial practices to enable open data may have positive upstream consequences. Dempsey and Teninbaum describe how machine-readable structuring of judicial opinions can improve efficiency in proceedings and the depth and quality of understanding of outcomes. First, it can improve accessibility, increasing "the ability of those governed by the decisions to understand and make predictions about new cases, as well as speed up research and lower legal costs." From a computational perspective, normalization at the data structuring level can unlock more complex, relational applications, making it easier to understand the ecosystem of legal concepts and principles.

This type of normalization at the systems level, in how court systems structure, report, and overall open data, enables comparative analysis. Advocates could reliably assess and compare the outcomes of different systems, shedding light on not only problems, but also on best practices. This can enhance knowledge-sharing across court systems and promote standardization. In addition to comparison, aggregation of state legal data in particular can form regional and national pictures of justice equity, legal service delivery, systems performance, and more.

Another opportunity is anti-corruption applications. Freedom of information and open government can reduce corruption when coupled with media and internet freedom. In this instance, administrative data about the institution may be particularly valuable to understanding processes and staffing. Caseload statistics or data on jurists from a database like the one created by the Free Law Project can track court performance and outcomes.

e. Boost public participation in law and policymaking ("CrowdLaw")

Originally proposed by Professor Beth Noveck in 2014,[14] CrowdLaw is "online public participation leveraging new technologies to tap into diverse sources of information, judgments and expertise at each stage of the law- and policy-making cycles to improve the quality as well as the legitimacy of resulting laws and policies."[15] In fact, there is promising evidence that CrowdLaw allows for rich, value-added public participation in law and policy making.

In 2012, CrowdLaw was used to collaboratively draft open data policies across the US. Those interested in contributing to New York City's 2012 Open Data Law could suggest feedback and edit prose on the draft. The City of Pittsburgh's mayor similarly opened up drafting of a local open data ordinance. This brought citizens into the lawmaking process, with the added benefit of exemplifying the open ethos. As laws are increasingly available in open and interactive formats, citizens can generate novel, impactful insights to improve the quality and relevance of the law.

Open legal data can help citizens to assess the implementation of laws and policies. When *I Quant NY* used parking ticket data to identify systematic ticketing of what were mostly legal spots, it became a prime example of the benefits of opening data. Ben Wellington identified a gap in the enforcement of

an obscure law that was proposed and adopted into traffic rules but not yet formally voted upon by the New York City Council.

3. Challenges and breakthroughs to open more legal data

A decade has passed since the Open Government movement took off in earnest, it has similarly been a decade of people calling for similar openness and innovation in the legal system. Fortunately, there is an engaged community of legal tech enthusiasts, creative access-to-justice advocates, academics promoting a culture of openness, and practitioners eager to modern technology and practices into the legal system. Even still, there is ample opportunity to open courts and legal institutions. With the above in mind, here are a few considerations for opening more and better legal data.

a. Law and policy frameworks

- Executive (i.e., governors, mayors) policies provide a foot in the door for opening data, typically within executive agencies, including Departments of Justice. From there, the legislature can enshrine the policy, and apply it to broader branches of government. Advocacy efforts that focus on executive public servants can usher in policy frameworks, to be tested and implemented before full enactment via the legislature.
- The UELMA does require a type of technical infrastructure nor a particular format for data, and is a lost opportunity for standardization and normalization. Shifting towards a unified data schema for materials covered by the UELMA could be a great boost for interoperability between distinct legal systems. A UELMA amendment could achieve this, or perhaps a new form of Model Act that applies to electronic representations of laws.

b. Meeting the open standard

- As the NCSC survey shows, many courts already have the capacity to make data available online already, it just might not be publicly accessible or free to use. In cases where the data is not free, it is often due to the use of a contractor. In these cases courts can partner with civil society organizations to enhance the openness of information while still reducing cost. This is a promising area of focus for advocacy efforts, and can be followed by capacity-building in less-digitized courts.
- Courts that already make data available can adopt bulk downloading. A single state made publicly-available legal data available for bulk download. Rather than a technical barrier, respondents cited a fear of misuse of the data. This concern is antithetical to the open movement and requires more investigation and treatment by legal and open advocates.
- As Data.Gov showed, it is not only a matter of opening data. Maintenance is particularly of concern in the law given the deprecation of information, and in general it signals strong open data practices. Moving towards a method employed by govinfo to continuously update information via effective

data management practices is key, particularly for older open data platforms that are past the initial push, and must now shift to maintenance.

c. Ecosystem-building

- **The richest and most exciting resources for legal data are not provided by public institutions, but generally come from scrappy efforts in civil society and private sectors. This affirms that there is a gap to fill.**
- Surveys of open data are becoming progressively outdated. The Open Knowledge Foundation's Global Data Index dates to 2017. The State Court Organization Survey used in this paper also dates to 2016-2017. The Court Statistics Project's most recent annual reporting is from 2018. As time passes, these resources become less and less useful. Fostering interest in these questions among students of law and policy can help build more and better tracking of open data and practices among local justice institutions.

Drawing on the advocacy of forward-thinking legal technologists like Jameson Dempsey, an open culture for law is one of the greatest opportunities in the coming decades for legal institutions and justice systems. As it stands, the law as a profession often seeks to close itself off, from costly and competitive legal education, to alienating citation mechanisms and jargon, to bar administration and membership. There is deep value in the professionalization of the law, but there is a trade-off in the approachability that need not apply to every aspect of the practice.

Openness of information is one step forward, and it is an important one. Not only does it provide the fuel for advanced computational applications, and the interrelationship of disparate legal systems, but it is also a chance to evolve the field. In a multifaceted country tackling problems of pandemic and systemic proportions, we are reminded that our justice system should embody transparency and openness, and integral to that is progressively opening troves of closed, digitized information. Information is truly power, and may mean enhanced access to justice, improved law and policymaking, advanced legal analytics, bolstered community activism, and, most likely, unimaginable gains for the very institutions that opened up in the first place.

Acknowledgements

Many thanks to individuals whose work and advocacy invited me to dig into this topic: Jameson Dempsey, Daniel Hoadley, Andy Silva, and to the teams doing the work at orgs like Measures for Justice and the Free Law Project. Added thanks to Bryan Wilson and Melis Emre.

Gabriella Capone currently supports the Reimagine New York Commission on its public engagement strategy. She tweets @thegabcap.

Footnotes

1. For global context, there are a number of barometers and indexes that look to compare the relative openness of countries around the world. The US ranked 11th in the most recent Global Open Data Index, with an estimated one-third of possible public datasets made open. ↩
2.
Open legal data is not a monolith, but meant to fulsomely encompass information from legal institutions. This piece focuses on such data from public institutions, to enable more open, computational, innovative, and equitable efforts. For a broader view of legal data, refer to Hagan, Dempsey, and Jímenez, "A Data Commons for Law" Medium, April 1, 2019. Available at: https://medium.com/legal-design-and-innovation/a-data-commons-for-law-60e4c4ad9340.

↩
3. *See* Open Government Initiative. Available at: https://obamawhitehouse.archives.gov/open; *See also* "The State of the Union of Open Data" Data Foundation, 2nd edition, February 2018, at: https://www.datafoundation.org/the-state-of-the-union-of-open-data-ed-2#government-data. ↩
4. Many open data stories begin with FOIA. *See* Beth Noveck, "Rights-Based and Tech-Driven: Open Data, Freedom of Information, and the Future of Government Transparency" 19 Yale Hum. Rts. & Dev. L.J. 1 (2017). Available at: https://heinonline.org/HOL/LandingPage?handle=hein.journals/yhurdvl19&div=3&id=&page= ↩
5. 17 U.S.C. § 105. ↩
6. *See* Project Open Data Dashboard, data.gov. Available at: https://labs.data.gov/dashboard/offices/2020-03-31/qa. ↩
7. The University of Chicago also hosts an excellent repository of judicial data. *See "Empirical Legal Research: Data Sources & Repositories"* The University of Chicago Library. Available at: https://guides.lib.uchicago.edu/c.php?g=747050&p=5352439. ↩
8. E.g.: case statistical information, administrative information such as workloads, information on the nature and status of proceedings. ↩
9. The judiciary's retention policy means that some information may no longer be accessible over time -- this merits future investigation. Timelines to enable effective data governance may impact information that should remain publicly-available over longer periods of time. *See* "Guide to

Judiciary Policy" Vol. 10, ch. 6. Available at: https://www.uscourts.gov/sites/default/files/vol10_ch6_appx_6b.pdf. ↩

10. *McBurney v. Young*, 569 U.S. 221 (2013). ↩

11. The survey includes the 50 US states, the District of Columbia, Puerto Rico, Guam, Northern Mariana Islands, and US Virgin Islands. American Samoa, tribal nations, and other sub-Federal jurisdictions are not reviewed and no doubt merit dedicated coverage. ↩

12. Young and Verhulst provide a framework that is a helpful reference, highlighting the power of open data to: improve government, empower citizens, create opportunity, and solve public problems. *See* "Open Data's Impact" The GovLab. Available at: https://odimpact.org/. ↩

13. Capone, Gabriella and Alsina, Victòria Burgués. "Enhancing Public Governance Quality and Effectiveness through Open Innovation." 2019 Transatlantic Dialogue. ↩

14.
Noveck, B. (2018). 'Crowdlaw: Collective Intelligence and Lawmaking,' *Analyse & Kritik*, 40(2): 359–380.

↩

15. Alsina, V., and Martí, J.L. (2018). 'The Birth of the CrowdLaw Movement: Tech-Based Citizen Participation, Legitimacy and the Quality of Lawmaking,' *Analyse & Kritik*, 40(2): 337–358. ↩