Received: 4 December 2020 | Accepted: 15 February 2021

DOI: 10.1002/aaai.12039

SPECIAL TOPIC ARTICLE

The Promise of AI in an Open Justice System

Adam R Pah[1] | **David L Schwartz**[2] | **Sarath Sanga**[2] | **Charlotte S Alexander**[3] | **Kristian J Hammond**[4] | **Luís A.N. Amaral**[4] | **SCALES OKN Consortium**[5]

[1]Kellogg School of Management, Northwestern University, Evanston, Illinois, USA

[2]Pritzker School of Law, Northwestern University, Chicago, Illinois, USA

[3]Robinson College of Business, Georgia State University, Atlanta, Georgia, USA

[4]McCormick School of Engineering, Northwestern University, Evanston, Illinois, USA

[5]Northwestern University, Evanston, Illinois, USA

Correspondence
Adam R Pah, Kellogg School of Management, Northwestern University, Evanston, IL, USA.
Email: a-pah@kellogg.northwestern.edu

Funding information
National Science Foundation Convergence Accelerator Program, Grant/Award Numbers: 1937123, 2033604

Abstract

To craft effective public policy, modern governments must gather and analyze data on both the performance of their public functions and the responses by the public. Federal administrative agencies such as the Patent Office and Centers for Disease Control routinely do this, as does the United States Congress. More importantly, they make such data freely accessible. Within the United States government, however, the judicial branch is a conspicuous outlier. In theory, federal court records could be used to evaluate the efficiency and fairness of the justice system. In practice, court records are effectively out of reach because they sit behind a government paywall. This financial barrier, along with an equally important myriad of technical obstacles, have forestalled the development of AI-driven analysis that could enable a systematic understanding and evaluation of the work of the courts.

The Systematic Content Analysis of Litigation EventS Open Knowledge Network (SCALES OKN) seeks to address this situation by transforming the transparency and accessibility of court records. The SCALES OKN will potentiate the development of new AI solutions that will benefit the judiciary, legal scholars, and the public. In this article, we outline some of key financial, technical, and policy challenges to developing novel AI solutions.

INTRODUCTION

The judiciary makes only a fraction of the data it collects freely available to the public ("Federal Judicial Caseload Statistics", 2018). Most federal court records cost 10 cents per printed page to view online ("PACER User Manual for CM/ECF Courts", 2019). The judiciary thus stands as an outlier as a branch of the federal government that charges the public to access public records. The executive and legislative branches, by contrast, produce and release data on their performance—Congressional votes, Food and Drug Administration reports, White House press briefings, and much more.

The purpose of the law that authorized the federal courts to charge such fees was to compensate courts and clerks for the time and resources needed to physically print or photocopy pages. The true cost of access plummeted after federal court records were moved online, yet the charges continued. Without access to federal court records, simple questions like "What percent of cases settle?" or "How many cases involve local governments?" are essentially impossible to answer with any

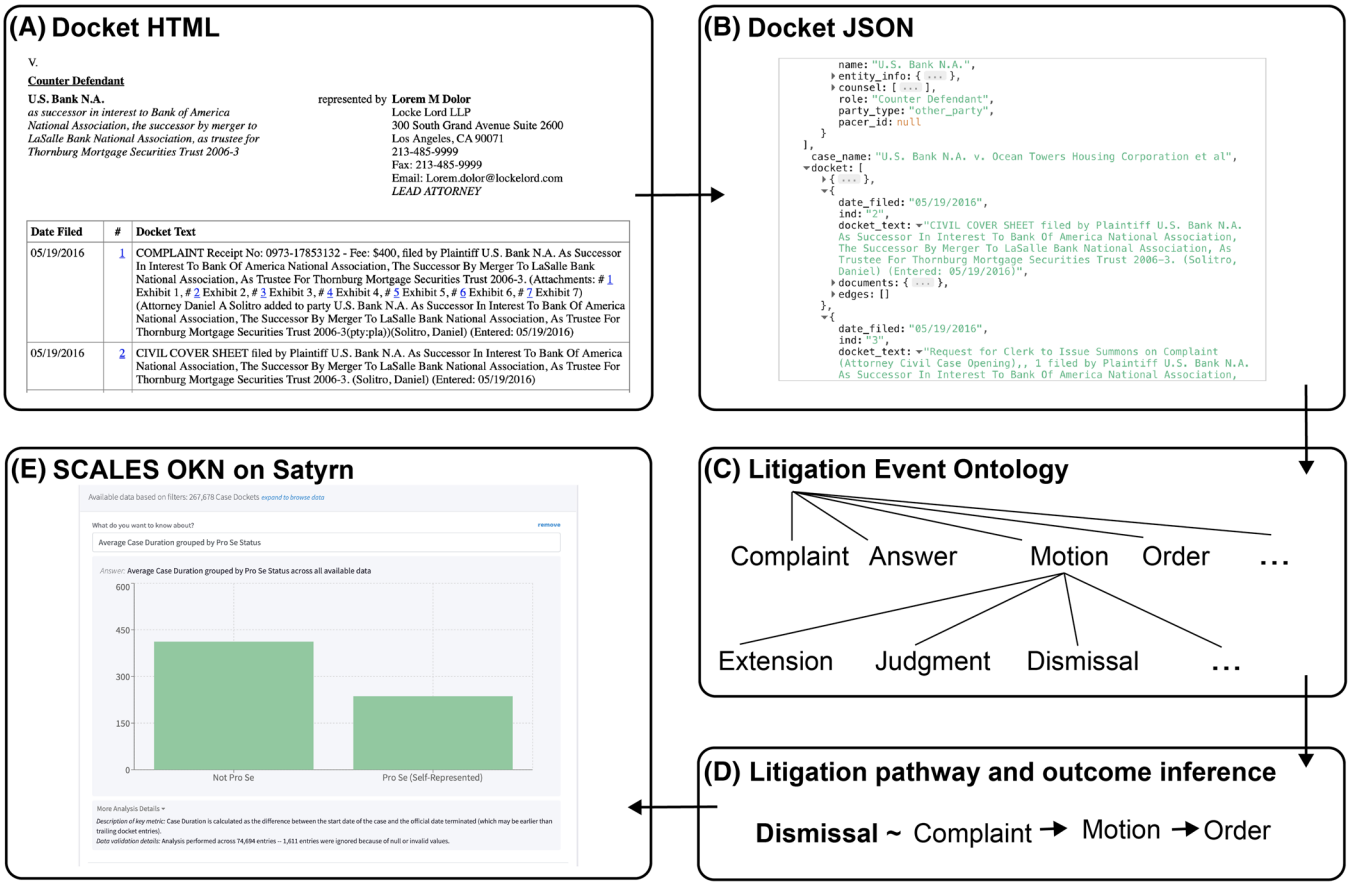

FIGURE 1 A court docket contains both semi-structured and unstructured text. The raw data are obtained as HTML (A) from the courts and extracted to a JSON format that enables further computation (B). Entities involved with a case—parties, lawyers, and judges—must be disambiguated to link across cases and to outside data. (C) The event ontology is used with NLP models to classify what the docket event is, this allows for (D) more complex models that can use this ontology to infer outcomes and the litigation path the case follows. Satyrn has a data-aware configuration, with knowledge of what types of analyses are supported for certain types of data, that populates the available analyses (E). Through the combination of the raw data, ontology, and data annotations, it can answer questions with visualizations (shown) or natural language

precision (Baude, Chilton, and Malani, 2017; Bielen et al., 2021).

Scholars and observers of the courts have long called for increased openness and experimentation in the courts to both study and improve the justice system (Lynch, Greiner, and Cohen, 2020). Indeed, federal courts are uniquely situated to do this. Each of the 94 districts courts could, in theory, experiment with its own local procedures and forms (Hammond, 2019) and recent work has demonstrated the potential for such experimental approaches (Pah et al., 2020). However, a necessary prerequisite for any such experiment is the ability to consistently document and measure the work of the courts—which is absent for most litigation events currently.

TECHNOLOGICAL APPROACH

Current court recordkeeping practices produce rich but largely unstructured text data. These data fall into two main forms: (i) chronological entries on a case's docket sheet (Figure 1(A)) and (ii) lengthy party- and judge-written documents filed in the case. The docket entries in a docket sheet maintain the record of a case's events from beginning to end. These entries make visible the strategic choices that the parties make along the way, as well as the response of the judge to those actions. In the absence of a higher level synthesis of the history of the case, docket sheets become the sole source of such information.

The first challenge for AI is the lack of a computational understanding of the free text of docket entries. Building such an understanding is significantly complicated by the fact that each jurisdiction, judge, clerk, and party can have their own vernacular for how they document litigation events and case progression. This complexity is not surprising since PACER, which serves federal court data, was built to facilitate case management by the court, not to provide open access or help study the system. It follows that there is no ground truth dataset—whether it be for identifying entities that appear in a case or for the systematic

labeling of litigation events—to facilitate model building. Without supporting annotations, the value of the data is significantly reduced. Our approach to court records is to model this language and annotate the dockets, making it so that judges, parties, and events are identified, searchable, and analyzable.

As an example of our data approach, we examined how document sealing procedures work across the courts. The right to open court data also brings with it the duty to protect the privacy interests of litigants and third parties. The courts of the United States and the documents generated in litigation are, by custom and law, generally open to public view, but litigants can request that some documents be sealed from public access if they contain trade secret or other confidential information.

Parties may want to seal filings for reasons unrelated to the sensitivity of the underlying information though. For example, a company may believe that multiple organizations are infringing its patents, but at first, it files a complaint against only one organization. To maintain a competitive advantage against future defendants, that plaintiff may want to seal information related to the current case, such as the plaintiff's theory of the patent's scope. The defendant may be willing to permit the plaintiff to seal the information since the other potential infringers are competitors to the defendant. Thus, sealing benefits both the plaintiff and the defendant over other competitors, but the public is disadvantaged. The judge is supposed to consider the public interest before sealing materials but only hears from the litigants in the first case. This dynamic provides the potential for unwarranted sealing.

We developed a regular expression-based algorithm to identify the docket entries associated with motions to seal court documents to study this further in patent cases filed in 2016. After multiple expert review rounds of algorithmic labels, we reach a final, purpose-specific model that can produce initial insights. For example, in the Eastern District of Texas (which saw almost 40% of all patent cases in 2016) successful motions to seal produce about 34 sealed items on average. In the other 93 federal district courts, by contrast, successful motions produced only about one sealed item on average (SCALES OKN Consortium, 2021). We construct a number of these purpose-specific regular expression-based models, all built with expert review and hand-coding, to bootstrap our initial training datasets. This allows us to build a large training corpus for more advanced NLP models to dramatically scale what our language models learn and, ultimately, can annotate in the docket sheets.

The second challenge is to provide our core users (researchers, journalists, and policymakers) with the ability to analyze such complex data and annotations. To do this, we are leveraging the Satyrn platform (Paley et al., 2021), which has a model of interaction that integrates natural language querying and visualization in a notebook-style interface (Kluyver et al., 2016) that allows for analysis of external datasets. Satyrn makes it possible for nontechnical users to systematically explore, analyze, and visualize complex datasets in an intelligent manner. Key to this is the analytics engine of Satyrn, which builds the universe of possible analyses that it can conduct for the user based on a dataset configuration. The configuration defines the relevant fields for analyses and their relationships, which then enables the analytics engine to infer which operations (e.g., average, year-over-year analyses) are applicable. Satryn presents users with potential analyses in natural language, constructs and executes the appropriate database query and subsequent analysis, and presents the results as visualizations and natural language—significantly increasing the information a typical user is able to unlock from the data (Figure 1(E)). Importantly, Satryn is designed to be dataset- and domain-agnostic so users can import their own datasets and increase the available analysis space.

CHALLENGES AND RELATION TO AI

One of the core challenges in making court records easy to analyze lies in developing protocols to disambiguate entitie—litigants, lawyers, judges, third parties, and others—and discovering the complex relationships amongst them as a case proceeds. Systematic Content Analysis of Litigation EventS Open Knowledge Network (SCALES OKN) is tackling this challenge by developing entity disambiguation and event ontologies. Recent advances in deep learning and transformer models have produced notable increases in the performance of named entity recognition (Devlin et al., 2019; Montani et al., 2020; Wolf et al., 2020). Already we have trained an entity recognition model that can accurately recognize judges in docket text and built a disambiguation pipeline to map the judge entities to their official biographical record, which enriches the potential analyses available to users.

Court records produce many unique entity recognition challenges though. One such challenge is that an individual may be sued in different capacities, and more generally that one person may effectively stand in for several other persons or entities. For example, a 2016 civil rights case from Indiana lists current Transportation Secretary Peter Buttigieg as a defendant. That case, however, was not brought against the person Peter Buttigieg, but instead in his capacity as then-mayor of South Bend, Indiana. Knowing in what role a party relates to a case can dramatically alter the understanding of the case and its relevance as a feature in AI models. To help solve this and related issues, we are training models to predict what

class an entity is in a case (i.e., *foaf:Person*) and build disambiguation tools to link this party information with outside data, such as Securities and Exchange Commission filings and published patents, to create a richer data ecosystem. However, many challenges remain in this area and the work of the interested AI community will be critical for continued progress.

Similarly, answering even simple sounding questions such as "What fraction of cases filled in 2016 in the Northern District of Illinois settled?" is currently impossible to answer in an automated manner because we do not yet have a robust ontology and model of litigation events. Building such an ontology is not easy because docket entry text is not systematic; rarely is there a docket entry that simply states "This case has settled." We are actively developing litigation event ontologies and building classifiers that will label docket sheet entries according to the event to which they pertain. These models will transform the semi-structured text of case docket sheets into an easily understandable sequence of events and allow SCALES users to follow a case from beginning to end. Layering in case-level metadata such as claim type and party, judge, and lawyer characteristics will also allow exploration of trends in case paths and outcomes at scale.

We are presently building ontologies for events in both civil and criminal cases. For both, we are building a structure with three tiers of ontological labels. The first tier is the most general and represents the larger ontological litigation phases through which a case moves and that docket entries reflect, such as case opening, discovery, dispositive motions practice, and case closing (Figures 1(C) and 1(D)). The second tier captures the general classes of filings and events that compose a case: complaint, answer, motion, and order. The third tier is the most granular and differentiates among subtypes of the second tier. For example, in the third tier, we label party-filed motions as to their specific subtype, for example, motion to dismiss or motion for extension of time, and complaints as the first, second, or amended complaint.

We are building classification models that implement these ontologies and label docket entries appropriately for all three tiers. Importantly, while all docket entries will receive both a Tier II and Tier III label(s), not all docket entries will receive a Tier I label. As an example, a motion for additional pages filed by a lawyer seeking permission to submit an extra-long brief to the court would only be labeled a motion (Tier II) and a motion for additional pages (Tier III). However, it is not an important enough motion, in and of itself, to signal a larger litigation phase like our Tier 1 labels of discovery or case closing.

There are technological complexities in operationalizing the progress of a case into the understanding that an ontology provides. For example, in civil litigation, some Tier I labels like case beginning are more constrained and will attach directly to the "Complaint," "Writ," or "Notice of removal" Tier II labels, as these docket entries, in and of themselves, usually signal the beginning of a civil case. In contrast, a Tier I case ending label might attach to an array of docket entries, including a settlement, granted motion to dismiss, and trial verdict. Settlement, however, can appear on a docket sheet via many different constellations and combinations of Tier II and III labels. For instance, the parties may file a Motion (Tier II) for approval of settlement (Tier III), in response to which the court enters an Order/Opinion (Tier II) granting (Tier III) approval. Only by identifying that sequence of Tier II and III docket entries can we discern that the case has ended via settlement, the relevant Tier I litigation ontology event.

A sociotechnical issue the SCALES OKN faces is how higher-level analytical results can be communicated to end-users in a way that preserves both ease of understanding as well as data provenance and the associated transformations. Satyrn provides results in the form of both visualizations and natural language descriptions. Users can either ask follow-up questions, now in the context of the current results, or focus on aspects of those results through direct manipulation of the visualizations. However, while this higher-level presentation of the data preserves ease of understanding, it also necessitates additional work to communicate data provenance, data quality, and how it was manipulated to this final form. Satyrn has some features already to communicate to the user how the data has been manipulated based on their queries and analysis, for example the system displays the changing total number of cases as filters are added or removed. However, more complex data issues—that is, how many cases have closed will differ on a year-by-year basis based solely on filing recency—require more testing to identify robust solutions to communicate to the user what are the caveats related to the requested analysis.

CURRENT STATUS

The SCALES OKN data ecosystem and user experience are under active development, but we have hit notable benchmarks that pushed the OKN to a beta version that we will use to conduct long-term user testing. We have focused on accumulating a starting dataset that will support the development of both longitudinal and cross-district analyses. In the beta version of the SCALES OKN, we have records for all civil and criminal cases filed in all 94 federal district courts in 2016 to support cross-district analysis. We also have all cases filed in the Northern District of Illinois from 2007 to 2017 to support longitudinal analyses. In addition, we have implemented a crosswalk that allows the docket

reports to be merged with the Federal Judicial Center's Integrated Database, a standard administrative dataset in empirical legal analysis that contains some case-level data, so users can make comparisons between the administrative and raw data, as well as the Federal Judicial Center's Judge Biographical data.

Importantly, we have also integrated our judge entity recognition and disambiguation annotations into the beta version of the SCALES OKN. The judge who is in charge of a case can, and does, change as a case progresses due to a number of potential reasons (retirement, promotion, etc.). Judge attribution at the individual docket entry level is an essential ingredient to produce accurate and robust statistics when asking questions that concern judicial actions (i.e., granting motions) or court function (caseload).

FUTURE PLANS

In the next year, our plan is to continue enriching the data ecosystem and expand user testing of the system and its data modeling, with a planned full public release of the SCALES OKN on Satyrn and the underlying data in Summer 2022. We are targeting to have at least 2 years of records available in Satyrn, along with our model-based labeling of entities and litigation events and data crosswalks to other administrative data, for users to search and analyze in our public launch.

During this year, we also plan to continue one of our core activities—helping the courts with systematic analyses of how they currently function. Our goal is to both advocate for legislation to make record access free and continue to demonstrate the value of this data, to both the courts and the public, once it is transformed to information. Long-term, the goal of the SCALES OKN is to become the primary open access resource that supports systematic analysis of court records at both the state and federal levels. This will require significant continued investment across a number of dimensions (data acquisition, modeling, user experience, and community engagement), but the potential for research and improving the transparency and function of the courts for the public is tremendous.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation Convergence Accelerator Program under Grant Nos. 1937123 and 2033604.

CONFLICT OF INTEREST

The authors have no conflicts of interest to disclose.

REFERENCES

Baude, W., A. S. Chilton, and A. Malani. 2017. "Making Doctrinal Work More Rigorous: Lessons from Systematic Reviews." *The University of Chicago Law Review* 84(1): 37–58. https://www.jstor.org/stable/44211824

Bielen, A., D. Kwan, D. Luetger, N. Maki, V. McFadden, and M. Myers. 2021. "Administrative Office of the U.S. Courts CM/ECF Path Analysis." https://free.law/pdf/pacer-path-analysis.pdf

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." http://arxiv.org/abs/1810.04805

"Federal Judicial Caseload Statistics." 2018. https://www.uscourts.gov/statistics-reports/federal-judicial-caseload-statistics-2018

Hammond, A. "Pleading Poverty in Federal Court." *The Yale Law Journal* 128(6): 1478–564. https://papers.ssrn.com/abstract=3102522

Kluyver, T., B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, and Jupyter Development Team. 2016. "Jupyter Notebooks—A Publishing Format for Reproducible Computational Workflows." In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 87–90. Amsterdam: IOS Press. https://doi.org/10.3233/978-1-61499-649-1-87

Lynch, H. F., D. J. Greiner, and I. G. Cohen. "Overcoming Obstacles to Experiments in Legal Practice." *Science* 367(6482): 1078–80. https://doi.org/10.1126/science.aay3005

Montani, I., M. Honnibal, M. Honnibal, S. Van Landeghem, H. Peters, A. Boyd, M. Samsonov, J. Geovedi, J. Regan, G. Orosz, P.O. McCann, S.L. Kristiansen, D. Altinok, Roman, G. Howard, W. Phatthiyaphaibun, S. Bozek, Bot Explosion, B. Böing, M. Amery, L.U. Vogelsang, Y. Tamura, P.K. Tippa, J. Fukumaru, G. Dubbin, V. Mazaev, R. Balakrishnan, J.D. Møllerhøj, T. O, and A. Patel. 2020. "spaCy: Industrial-strength Natural Language Processing in Python." https://doi.org/10.5281/zenodo.1212303

"PACER User Manual for CM/ECF Courts." 2019. https://www.pacer.gov/documents/pacermanual.pdf

Pah, A. R., D. L. Schwartz, S. Sanga, Z. D. Clopton, P. DiCola, R. D. Mersey, C. S. Alexander, K. J. Hammond, and L. A. N. Amaral. 2020. "How to Build a More Open Justice System." *Science* 369(6500): 134–6. https://doi.org/10.1126/science.aba6914

Paley, A., A. L. Zhao, H. Pack, S. Servantez, R. F. Adler, M. Sterbentz, A. Pah, D. Schwartz, C. Barrie, A. Einarsson, and K. Hammond. 2021. "From Data to Information: Automating Data Science to Explore the U.S. Court System." In ICAIL '21: Eighteenth International Conference for Artificial Intelligence and Law. https://doi.org/10.1145/3462757.3466100

SCALES OKN Consortium. 2021. "scales-okn/Research-Materials: Sealing Analysis." https://doi.org/10.5281/zenodo.5056441

Wolf, T., L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush. 2020. "HuggingFace's Transformers: State-of-the-art Natural Language Processing." http://arxiv.org/abs/1910.03771

AUTHOR BIOGRAPHIES

Adam R. Pah is a Clinical Assistant Professor of Management & Organizations in the Kellogg School of Management at Northwestern University. His research focuses on the evolution of complex social systems.

David L. Schwartz is the Frederic P. Vose Professor of Law in the Pritzker School of Law at Northwestern University. His research focuses on empirical studies of patent law and judicial behavior.

Sarath Sanga is a Professor of Law in the Pritzker School of Law at Northwestern University. His research focuses on corporate law and contract theory.

Charlotte S. Alexander holds the Connie D. and Ken McDaniel WomenLead Chair as an Associate Professor of Law and Analytics in the J. Mack Robinson College of Business at Georgia State University. Her research focuses on uncovering patterns in civil litigation using text mining, natural language processing, and machine learning.

Kristian J. Hammond is the Bill and Cathy Osborn Professor of Computer Science in the McCormick School of Engineering at Northwestern University. His research focuses on developing human capabilities onto machines.

Luís A.N. Amaral is the Erastus Otis Haven Professor of Chemical and Biological Engineering in the McCormick School of Engineering at Northwestern University. His research focuses on the emergence, evolution, and stability of complex social and biological systems.

SCALES OKN Consortium includes the numerous professors, students, and staff members that contribute to the development of the SCALES OKN. The list of consortium members can be found at https://scales-okn.org/research/.

How to cite this article: Pah, A. R., Schwartz, D. L., Sanga, S., Alexander, C. S., Hammond, K. J., Amaral, L. A. N., and SCALES OKN Consortium. 2022. "The Third AI Summer." *AI Magazine* 43: 69–74. https://doi.org/10.1002/aaai.12039