Silicon Flatirons
KNOW WHAT'S NEXT.

Fall 2022 Technology and Legal Services Report

Harnessing the Power of AI and Data Organization to Better Utilize Institutional Knowledge and Data

Rapporteur: Megan Decker[1]

Table of Contents

[1] Third year law student, University of Colorado Law School. Published February 14, 2023

Abstract

A law firm may have mountains of data on past cases and transactions, yet no efficient way to mine or search that data for useful practice tips for future litigation and deals. Emerging artificial intelligence ("AI") and available data organization tools offer the promise of better access to the data which exists within an organization.

This Silicon Flatirons **Technology and Legal Services Report** examines emerging possibilities for institutions, especially legal organizations, to better leverage knowledge and data. The Report arises from a Roundtable discussion convened by Silicon Flatirons on September 27, 2022 (herein, the "Roundtable"). The Roundtable included entrepreneurs, legal tech consultants, law librarians, attorneys, investors, venture capitalists, legal outsourcing company representatives, and members of the academic community. The discussion was co-moderated by Jason Adaska, Director of Software Engineering and Innovation Lab at Holland & Hart LLP and Brad Bernthal, Associate Professor of Law at the University of Colorado Law School.

The Report captures three important insights:

- Mining market data with AI would reduce information asymmetry between parties in a transaction (i.e., it would help equalize what parties know);
- Law firms could better utilize their institutional knowledge by building an AI assisted, searchable repository of information; and
- another option for organizations, where building an in-house solution is unworkable, is to invest in fledgling tech companies or to tap a tech company to build a data organization solution.

Following the Introduction below, this Report proceeds in three Parts which describe how artificial intelligence ("AI") and data organization tools can (1) balance information asymmetries in relationships and transactions, (2) enhance an organization's access to its institutional knowledge, and (3) aid organizations in determining overarching business strategy.

Introduction

Imagine visiting the Library of Congress, the largest library in the world with over 173 million[2] items in the collection including The Gutenberg Bible, Thomas Paine's *Common Sense*, and Mark Twain's *Adventures of Huckleberry Finn*,[3] and not being able to locate the relevant book you are seeking amid the overwhelming available resources. Imagine knowing that all this wisdom, insight, and expertise is just out of your grasp, and not having a way to access it.

This, in a nutshell, is the challenge organizations face today with institutional knowledge and unstructured, disorganized data.

Institutional knowledge is an organization's collective memory. "It encompasses all the job-related facts and information that live in each individual employee's head."[4] Converting this individual knowledge into a shared institutional memory is critical for an organization's growth and development. If institutional knowledge is allowed to leave with each departing employee before it is captured by the organization, "it slows productivity and creates confusion and miscommunication that could have dramatic repercussions for the [organization]."[5]

Capturing this institutional knowledge is not as simple, however, as having an exit interview with a departing employee or having the employee return their work laptop. Technology, like a work laptop that can store all an employee's work-related material, promotes capture of institutional knowledge (at least to a degree). But it presents organizations with a different problem: if the data on the laptop is unstructured or disorganized, how useful is it to the organization?

Now take that one employee laptop, multiply it by one thousand, and you start to get the volume of data organizations have access to but may not be able to use because it is so unstructured or disorganized. Notably, this problem is not exclusive to passing institutional knowledge from departing employees to new employees. Information asymmetry and disorganized data exist across several transactions and numerous relationships.

---

[2] Library of Congress General Information, LIBRARY OF CONGRESS https://www.loc.gov/about/general-information/#year-at-a-glance (last visited Oct. 16, 2022).

[3] Jennifer Gavin, *Library of Congress: Books That Shaped America*, LIBRARY OF CONGRESS (January 22, 2013), https://www.loc.gov/item/prn-13-005/.

[4] Joei Chan, *Why Not Sharing Institutional Knowledge is Costing Your Company Money*, 360LEARNING, https://360learning.com/blog/institutional-knowledge/ (last visited Oct. 16, 2022).

[5] *Id.*

Legal institutional knowledge in particular has challenges. Not only does it require domain expertise, but there is frequently a gap between plain language and "legalese" or legal writing that can be a barrier to efficiently sharing institutional knowledge. An experienced investor investing in a small startup has an information advantage, a company with a patent on a particular product or process knows they have a patent, but what that patent covers *exactly* may be trapped in their patent attorney's head or buried in their company information repository where they keep a copy of their patent.

Part I: Information Asymmetries in Relationships and Transactions

The term "information asymmetry" refers to an imbalance between relevant information known by one party but not another. A high degree of information asymmetry impedes efficient transactions. To illustrate how information asymmetry issues can frustrate a transaction, Otto Hanson, founder of TermScout,[6] introduced two types of transactions: buying Levi jeans and buying an artisan rug.

The first transaction – a jeans purchase - is almost effortless for the parties and involves no bargaining. Both parties have the same information, know a reasonable price for the jeans, and price is the only variable on the table. A customer typically purchases jeans without much – if any - haggling or wasted motion.

The second transaction, buying an artisan rug, is difficult for the parties. An initial problem is what constitutes a fair price. The seller, assuming she is active in the rug business, likely has far more information than the buyer about "market" price. The seller insists the rug is worth five hundred dollars. In turn, the buyer is concerned that the seller is trying to take advantage of her. She counters that she only has fifty and she wouldn't spend a penny more on a small entry way rug. Haggling ensues. Perhaps the parties settle at an arbitrary amount. Perhaps negotiations break down. In any event, the information asymmetry – and lack of a well known "fair" price – contributes significant friction to the transaction. And this is a simplified example where price is the only variable on the table. Complexity grows as other factors (e.g., quality of the rug, uniqueness of design, etc.) are involved.

A software licensing deal, Hanson points out, requires the negotiation of points which include not only price, but also the parties' liabilities, indemnification, data security, data usage, and so on. This is a high friction transaction. The parties have differing information and must negotiate several deal elements. You can probably imagine this negotiation going the same way as the artisan rug negotiation – e.g., each side takes a position, argues for it, and it becomes a time-consuming contest of wills and concessions to reach a compromise.[7]

On the other end of the spectrum, you may see what Hanson saw a few years back when he was practicing law. A client told him they wanted "to make their standard contract as aggressive as humanly possible

[6] Learn more about TermScout at http://www.termscout.com/

[7] *See* Roger Fisher & William Ury, GETTING TO YES 3 (1981).

because his clients were accepting it without reading it."[8] In transactions where neither party has insight into what's market or one party has an information advantage, we see inefficient, time-consuming negotiations, parties taking advantage of one another, and strained relationships.

To overcome this challenge, parties needed to be informed, but doing due diligence to become informed can take almost as long as the negotiation. To facilitate informed, efficient contracting, TermScout identified three things both parties needed: (1) transparency into what a contract says, (2) a shared understanding of what's "market" through market data, and (3) reasonableness: a contract cannot be one sided.[9] TermScout provides these three things in the software licensing space by gathering thousands of contracts from SaaS providers. Then through a combination of AI and human contract reviewers, TermScout will comb through these contracts to determine what's market. Finally, TermScout customers will upload their own or a third party's contract to TermScout and TermScout will benchmark the contract as balanced, vendor-favorable, or buyer-favorable depending on the presence of certain terms, conditions, and clauses previously identified by its AI and human contract reviewers.[10]

Others have noticed the challenges information asymmetry presents in transactions, and where TermScout is focused on software-licensing companies like Lexis Nexus,[11] Westlaw,[12] and Bloomberg[13] each have an AI powered tool for determining what's market in the mergers and acquisitions, antitrust, finance, labor and employment, intellectual property, and other spaces ("Deal Analyzing Tools"). These Deal Analyzing Tools are limited in that they can only capture terms from public deals because they source their information from the Security

---

[8] Otto Hanson, (Otto Miguel Hanson), LINKEDIN (May 2022), https://www.linkedin.com/feed/update/urn:li:activity:6928023965848276992/.

[9] TERMSCOUT, ABOUT US https://www.termscout.com/about (last visited Oct. 16, 2022).

[10] TERMSCOUT, HOW IT WORKS https://www.termscout.com/how-it-works (last visited Oct. 16, 2022).

[11] Lexis+, *Market Standards – M&A*, LEXIS NEXUS https://plus.lexis.com/practice-advisor-market-standards-search/corporate-mergers-acquisitions?pdmfid=1530671&crid=eca817f8-3828-4308-b0dd-924792a2defe (last visited Oct. 16, 2022).

[12]Thompson Reuters, *Practical Law: What's Market*, WESTLAW https://1.next.westlaw.com/Browse/Home/PracticalLaw/WhatsMarket?transitionType=Default&contextData=(sc.Default)&navId=A69756D811B60C80367051FA6ED84716 (last visited Oct. 16, 2022).

[13] BLOOMBERG LAW, https://pro.bloomberglaw.com/draft-analyzer/ (last visited Oct. 16, 2022).

Exchange Commission's (SEC) Electronic Data Gathering, Analysis, and Retrieval system (EDGAR).[14]

All these tools present multiple advantages to transacting parties. William Ury and Roger Fisher, participants of the Harvard Negotiation Project and authors of *Getting to Yes* postulate that negotiations should end with wise agreements or agreements that "meet the legitimate interests of each side to the extent possible, resolves conflicting interests fairly, is durable, and takes community interests into account."[15] Part of a wise agreement is taking time to "invent multiple options looking for mutual gains before deciding what to do"[16] and "insisting that the result be based on some objective standard."[17] Roundtable participants suggested that tools such as TermScout can help parties invent options and base their agreements on objective standards by shedding light on what other comparable parties are agreeing to in their transactions. If parties are stuck on a particular point, they may turn to Deal Analyzing Tools to gather information about what other parties are contracting and draw ideas from their terms. Additionally, a seller will be able to point to a contract run through TermScout which received a "balanced" score and have a defensible argument why their contract is fair to both parties and on par with what other sellers are doing thus enhancing the buyer's feeling of being treated fairly as well as knowing she's getting a better, or at least comparable, deal to a deal she could make with a different seller. Finally, the buyer knowing she's getting a fair deal, is more likely to return to this seller and engage in another transaction thus strengthening the relationship between the parties.

Although market data mined with AI can help to balance the information asymmetry between parties, it doesn't mitigate all power imbalances between parties. Roundtable participants identified several other factors that influence market: relative power between parties, term of relationship (one-off short term contract vs repeat contracts over several years), and course of dealing in a particular field can all impact what is considered market in a transaction. Roundtable participants also noted that market may not be established in new

---

[14] *See* Thomas Reuters, *Practical Law: Public Merger Agreements*, WESTLAW https://1.next.westlaw.com/Browse/Home/PracticalLaw/WhatsMarket/PublicMergerAgreements?transitionType=Default&contextData=(sc.Default)&navId=B4C1E29A3F14ABF6E015F41291F3BF15# (last visited Oct. 16, 2022) ("Coverage: The Public Merger Agreements database includes all acquisitions of US reporting companies (excluding REITs and debt-only issuers) with a signing value of at least $100 million since January 1, 2009, as well as all deals in 2008 with a signing value of at least $250 million and a selection of deals signed in 2007.").

[15] Roger Fisher & William Ury, GETTING TO YES 4 (1981).

[16] Roger Fisher & William Ury, GETTING TO YES 11 (1981).

[17] *Id.*

industries such as privacy where terms are volatile due to an ever-changing regulatory space. Hanson noted that some of these concerns would be solved with time and others with advances in Data Analyzing Tools that allow users to input their situational factors such as relationship of parties into the tool.

Part II: Granting Widespread Institutional Access to Institutional Information

Curating and preserving institutional knowledge is not a new concept. Just google how to preserve institutional knowledge and you get more than sixty million results which, surprisingly, remain relevant up to the twentieth search result page. Organizations have been trying for decades to ensure they have a process for capturing, preserving, and sharing institutional knowledge. The advantages of preserving institutional knowledge are obvious: it saves organizations time and money. A Roundtable participant says as a new attorney, she would love to have firm data in a place she could access it rather than having to ask the partners at her firm and take them away from their work. Another motivating driver, at least for law firms, is customers who, in an effort to avoid being a cost-center to their organization are performing a cost-benefit analysis and then negotiating for fee caps with their law firms based on their analyses.[18] This means law firms must become as efficient as possible in order to stay profitable. However, off-the-shelf data retrieval comes at a big price and even then, its generic data retrieval. Rather than purchase off-the-shelf solutions, large law firms such as Holland & Hart have turned to building their own in-house solutions.

Gareth Middleton and Jason Adaska are the Principal Data Scientist and Director of Innovation Lab at Holland & Hart, respectively, and are working to build a large-scale model powered by AI to capture institutional knowledge and enable attorneys to better utilize the data that law firms gather as a by-product of practicing law. To illustrate, Middleton describes a better search tool for the firm's litigators. Instead of searching a document by title or keyword, a litigator may want to search for documents within a certain time period, or documents produced by a specific attorney, or heard before a specific court. All this information is contained in a caption which appears on a cover sheet attached to all briefs filed in court. Using data gathered from Holland & Hart, Middleton trains a model to recognize specific data in a caption on a cover sheet. And because this model was trained using Holland & Hart data, Middleton explains, it's specific to the way Holland & Hart formats it's captions.

Middleton also highlights that feeding the model training data can be difficult because the training data usually needs to be provided by people. For example, attorneys or other personnel familiar with a cover sheet need to identify valuable information for the model to learn from and need to do this millions of times over for the model to have

[18] Harvard University, *Legal Tech and Law Firms In-house*, YOUTUBE (Oct. 14, 2019), https://www.youtube.com/watch?v=cAsDBaAULGo.

enough training data to be useful. However, Holland & Hart instead writes rules to generate training data to sidestep the time-consuming and expensive process of having people provide the training data. This works because cover sheets are relatively standard. Middleton and his team examine a few captions then write a rule that states, in layman's terms, if the computer sees a name, followed by an initial and then another name, that's probably the attorney. These rules then generate training data for the model. Now Middleton and his team take all the captions that exist at Holland & Hart, send them through the set of rules and produce a training set that can train the model. This can be done in a matter of minutes whereas it would have taken a team of staff hours or days to construct the same training data set.

Having this model allows litigators to quickly find all the cases filed by a specific attorney. Imagine being a new attorney and wanting to know how the partner you're reporting to likes to write briefs – run a search through the model with her name and suddenly you have all the briefs she's ever filed that you can read and use as examples for your own brief. You as a new attorney just saved your partner and your firm time and money that they would have spent training you. Or imagine wanting to know how opposing counsel structures their arguments or how a certain judge usually rules. Again, search the attorney or the judge's name and you have a litany of resources you can use to prepare your case. This not only saves you time, but also ensures that your clients are getting the best representation.

Middleton's example is only one model and application for utilizing data gathered as a by-product of practicing law, and Holland & Hart is not the only firm that has recognized the value of easily searchable information. Bryan McCutcheon is a Knowledge Management Attorney for Gibson Dunn & Crutcher. He works with the firm's leadership to "improve efficiency by optimizing technology solutions and improving workflow processes."[19] McCutcheon works on the transactional side helping his firm develop something like the Data Analyzing Tools examined in Part 1 of this Article. However, these tools are specific to deals done by his firm.

Roundtable Participants also noted two areas for further research regarding in-house data retrieval tools are integrating natural language processing into an in-house data retrieval tool and exploring data solutions for *capturing* institutional knowledge.

By building an AI assisted, searchable repository of information, firms like Holland & Hart and Gibson Dunn do not have to start from zero with each new transaction or litigation. Instead, they use institutional knowledge to move parts of the undertaking from both into repeatable

[19] Bryan McCutcheon, (Bryan McCutcheon), LINKEDIN (Oct. 2022), https://www.linkedin.com/in/bryan-mccutcheon-416a114/.

processes which in turn leverages economies of scale thus saving the firm and, more importantly, the client time and money.

Part III: Aiding Organizations in Overall Business Strategy

Of course, it would be remiss to not mention that building an in-house tech solution has its own expenses, and it's important for an organization to determine whether to buy an off-the-shelf solution; build their own; or use a hybrid solution by tapping a tech company to build something for it, modifying an existing solution, or building through investment (i.e. investing in fledgling tech companies).[20] NLPatent is an excellent example of a tech startup designed specially by patent attorneys to take the mystery out of patents. For most tech companies, a large part of their value is in their intellectual property. Protecting that intellectual property, therefore, becomes of the utmost importance. A significant part of protecting intellectual property is protecting an organization's patents.[21] When it comes to patents, however, knowing exactly what the patent covers or what someone else's patent covers can be difficult. Patent language is notoriously obtuse and technology specific, and NLPatent bridges the gap between what the patent says and understanding what it means.[22]

NLPatent uses a modern Natural Language Processing model to specifically understand patents. Users input full sentence queries, the AI understands the context, and results are generated; these results can further be refined based on user input. For example, the model understands the word "bank" when someone says, "I need to go make a deposit at the bank" is different than when someone says, "Let's have lunch by the bank of the river." Stephanie Curcio, CEO and co-founder of NLPatent, demonstrates how this technology can help companies with an example. She runs the phrase, "A robotic hoover than can automatically navigate a room in a house. It uses navigation software to avoid bumping into things. It can be scheduled to run at a certain time of day," through NLPatent's model. The model comes back with several patents for autonomous vacuum cleaners despite vacuum cleaner being nowhere in the description.

The Roundtable participants also discussed how beneficial it would be to combine NLPatent and TermScout, which would allow companies to use a natural language search to analyze contracts as well as patents. Additionally, participants noted that Google has an existing patent

---

[20] Aebra Coe, *Build versus Buy: 5 Ways BigLaw is Tackling Legal Tech*, LAW360 (2018), http://s3.amazonaws.com/cdn.orrick.com/files/Build-Versus-Buy-5-Ways-BigLaw-Is-Tackling-Legal-Tech.pdf

[21] Thomas Alsop, *Companies with the most U.S. patents granted to them in 2020 and 2021*, STATISTA (July, 1, 2022), https://www.statista.com/statistics/274825/companies-with-the-most-assigned-patents/ (showing International Business Machines Corp (IBM) was issued 8,682 patents in 2021 and Microsoft Technology Licensing LLC was issued 2,418 in 2021).

[22] NLPATENT, https://www.nlpatent.com/about (last visited Oct. 16, 2022).

searching tool, but it was noted that some companies block Google entirely out of concern that Google would have undue insights into what they are considering patenting.

Through this technology, organizations can understand their own patent portfolios. They can copy-and-paste full paragraphs from issued patents or invention disclosures into NLPatent to quickly assess a patent's strength or determine patentability. The organization can use the tool to gain transparency into what it owns, what it's monetizing, and where it might be better suited to spend its resources. Having this kind of transparency into an organization's data will help it make strategic business decisions like when to go after someone who may be infringing on its patent, what ideas to put time and resources behind, or what ideas to abandon because an NLPatent search revealed that a prohibitively similar patent already exists. The organization need only to input what the competitor's technology does or a description of the new idea into NLPatent then see if the company's or another's patent comes back in the search results.

Conclusion

Organizations not utilizing institutional knowledge are wasting time and money but usually the institutional knowledge is inaccessible or unusable because it is unstructured or disorganized. Resources such as Deal Analyzing Tools, TermScout, NLPatent, and in-house AI models can help structure data into a useable format. Structuring data helps organizations overcome information asymmetries, become more efficient by transforming projects into repeatable processes and leveraging economies of scale, and aids organizations in making better decisions regarding overall strategy. In highlighting three solutions, the Roundtable illuminates possibilities for leaders and organizations to combat the challenge of organizing and centralizing institutional knowledge and data. Although the Roundtable discussion focuses on the intersection between AI and legal, the ideas and technology discussed works for any organization. In discussing these three solutions, this Article does not attempt to be comprehensive but to serve as a starting point for organizations looking to capitalize on their institutional knowledge and data.

About Silicon Flatirons Center

Mission

Silicon Flatirons' mission is to elevate the debate surrounding technology policy issues; support and enable entrepreneurship in the technology community; and inspire, prepare, and place students in these important areas. Learn more at siliconflatirons.org/about-us/.

Our Team

For more information about center leadership, faculty, staff, fellows, and advisory board, visit siliconflatirons.org/about-us/our-team/.

Our Supporters

Silicon Flatirons exists thanks to the generosity of our supporters and the strength of our community. We rely on their contributions to advance our mission to catalyze policymaking and innovation and to develop the next generation of tech lawyers, policy experts, and entrepreneurs.

For more information on current official Silicon Flatirons Supporters, visit siliconflatirons.org/about-us/supporters/.

Publications

We promote thought leadership and intellectually honest discourse not only in our events, but in publications from our team, our roundtables, and scholars presenting at our conferences. See more at siliconflatirons.org/publications/.