This first program in the Legal Analytics Series - Organizing Useful Practice Data - First Steps for Everyone focuses on teaching attorneys the basics of creating and organizing useful practice data. Many lawyers haven’t had real exposure to the idea of using data to drive decisions. They might not understand the idea, or they might not know how to start. In this respect, the legal industry is well behind other industries. Almost all of the principles that guide the process of using data to make decisions come from other industries – but they are easily adapted to legal practice. In addition, data-driven decision making is equally applicable to in-house attorneys or legal operations.

In “What is Data Driven Law,” Mary Juetten provides a useful article on data-driven law, describing the methods attorneys should use to bring a data-driven approach to their practice. She outlines the basic concepts:

- Start with the end in mind: identify what you as an attorney want to accomplish, find your pain points, and determine what changes you want to see in your practice.
- Process before Purchase; Data before Decision: determine what you want to accomplish and how you might achieve that goal before you start investing in new systems and technology. Obtain data, and use that data to make your decisions.
- Capture data near the source: capture your information as close to the beginning of the process as possible, and continue to collect it as the process continues.

The Elements of a Legal Data Strategy, provided by LexPredict, outlines the process of developing a legal data strategy. A data strategy consists of two high-level core concepts. The first is a top-down mission statement acknowledging the value of an organization’s data. The second is a framework for developing new data-related capabilities. The article steps through the four aspects of managing legal data: recognition, storage, publication and accessibility, and explains how to tie the four aspects together.

Kira Systems provides a concrete example of what legal data analytics can accomplish. Their white paper on Materiality Scrape provisions in merger and acquisition deals shows how advanced analytics can extract data from document sets and provide valuable insights. In the paper, they extracted out and analyzed the materiality scrape provisions from 89 private target merger agreements completed in 2017 and 2018. Among the insights:

- 81% of agreements in 2017, and 90% of agreements in 2018 included materiality scrape provisions.
- In 2017 42% of the provisions were limited to calculation of damages/losses only, while the number of limited provisions declined to 26% in 2018.
Program: Legal Analytics Series: Organizing Useful Practical Data: First Steps for Everyone

Presented by Legal Analytics Committee
Co-sponsored by Career and Practice Development; Professional Responsibility; Young Lawyers (Part of Young Lawyers Track)
Abstract

This program focuses on organizing useful practice data. Speakers will address issues regarding why collecting and organizing data is helpful to the legal practice, what the basics of organizing and structuring legal data are, and what products and systems can help lawyers organize practice data.

Table of contents

KPI Workbook Small Law Firm KPIs --- Mary Juetten
What Is Data Driven Law--- Mary Juetten
Identifying Your Ideal Clients （Link）--- Mary Juetten
LEXPREDICT the Elements of a Legal Strategy --- Warren Agin
LEXPREDICT Assessing Legal Data Strategy Maturity --- Warren Agin

Participant Bios
- Moderator: Juliet Moringiello, Widener University Commonwealth Law School
- Chair: Warren Agin, LexPredict, an Elevate Company
- Speakers: Mary Juetten, Traklight; Anne McNulty, Kira Systems
- Material Coordinator: Jiaying Christine Jiang, Emory Law
Get started with Data-Driven Legal Services

We hear much about data and analytics these days, along with lean and design thinking, but many lawyers either do not understand what is meant by the term data-driven, nor know where to start. Almost all of the principles that underpin using data to make decisions are from other industries and are in line with running a law practice as a business. This translates into welcome precedent for using data within legal services. In addition, data-driven decision making is equally applicable to in-house attorneys or legal operations.

What is Data-Driven Law?

In a nutshell, data-driven law means gathering information in the form of numbers, qualitative information, and calculations, both internally and from outside the firm. The data or measurements are compared to targets and appropriate adjustments are made where possible based on the numbers or information, not gut-feel. A data-driven approach encompasses using metrics, or key performance indicators (KPIs).

In speaking with lawyers about data and KPIs over the past almost five years, confusion clouds the topic, starting with the fact that all these terms mean basically the same thing. In other words, data-driven law is just using metrics (and data) to measure performance and provide indications for change or improvement.

Below are some of the myths that seem to surround data and metrics:

- **Law is different from other professions.** Simply, false. Having worked with lawyers, engineers, doctors, and as an accountant, I can tell you that the processes and work flows are similar enough to warrant using the same framework and approach.
- **Make it up as you go.** Some feel that experimentation and a fail fast approach means that data collection and measurement is not planned. In fact, like many other management tasks, the planning is the majority of the task. Identifying the correct data, the best way to capture that data in a cost effective manner, and measuring at the start are all critical to success.
- **Perfection is Required.** Collecting data allows for informed decisions but the process or system that is being measured or piloted will never be perfect. Those who wait too long, will miss opportunities and are not devoting enough time to experimentation. The old adage, better done than perfect, applies here.

Start with the End in Mind

A standard approach is recommended as the diagram below outlines. The first step calls for goal identification which can be narrowly done by one attorney or broadly developed as a mission, vision, and values exercise for a firm of any size. The focus for all businesses, including professions like the law, must be on clients, cash, and compensation.
Instead of trying to measure and fix everything at once, list all the pain points, problems or challenges based on each of the following: clients, cash, and compensation. Check if there is data to back up your assertions and decide which process impacts all of the above. Again, data is required to measure any changes and impact as you experiment with solutions.

Reprinted with Author’s Permission, copyright Mary Juetten, 2018.

Process before Purchase; Data before Decision

The work flows are similar for most service businesses or professions. In the diagram above, the processes are represented at a high level in chronological order. To start, after successful marketing, the client is either new or has a new matter. Once engaged, the client work product is completed, the bill sent and the matter closed.

In order to decide what data to gather, it’s important to map out the process for two reasons. First, there may be opportunities to spot inefficiencies or areas where technology or a different approach may improve the process. Second, a visual of the process makes identifying the
appropriate data for collection much easier. Change and technology should only be considered after this type of process review and with data to back up any purchase decision.

Data

Some attorneys have said that they do not have the data to measure or that they do not have time to go back for an entire year to find some. However in reality, it’s more about not understanding where to look and focus.

The chart below outlines the types of data, on the left hand side, that can be collected for each of the five high-level processes. Most of these are simple data points to start recording and you do not need a full year’s worth of data to make interpretations.

However, a key principle is to capture data as close to the beginning of the process as possible and then along the way. For example, the number of new matters is a great data point but it is also important to know how many consultations and visits or website inquiries and phone calls happened to generate that new work. Plus, in order to evaluate marketing effectiveness, the source of those clients should also be recorded.

Reprinted with Author's Permission, copyright Mary Juetten, 2018.
Conversion Rate Example

Building upon the above diagram, let’s examine the data for new clients. For example, in the last month a fictitious firm has 20 new potential client contact the firm as follows:

- 5 calls to the office main number;
- 12 website contacts; and
- 3 calls directly to attorneys.

The above is good information but it’s not complete. For all 20, knowing what prompted them to reach out to the firm is just as important as how they contacted the firm. Whether it was a call to the main line or an attorney, the following are good data points as far as source (and not an exhaustive list):  

How did you hear about our firm?  
- Former client  
- From a friend or colleague who is a former client  
- Social media  
- Advertising  
- Google search  
- Attorney speaking or writing

A website contact form can have the above as a question on a dropdown menu. It would be helpful to know that of the 20 potential clients: 15 came from social media versus from general google search or advertising; or 10 came from attorney speaking or 15 came back as former client. Therefore, it’s important to discover how each person contacted the firm. And, we have just started because we are only reviewing the potential clients.

We then can track any of those potential clients as they convert into actual clients, noting the source. With that data, we can evaluate our marketing efforts and more. For example, if from the 20, there were 5 new matters, that would mean a 25% conversion rate. We can then track those 5 new matters in terms of revenue, profitability, and client satisfaction.

We would also wish to look back to see where the 5 originally came from, let’s say all came as phone calls to the main number and none of the other 15 from website or attorney turned into any actual work. That could mean that perhaps the messaging on the website and in speeches might not be attracting the right potential clients. The data is knowledge and therefore, powerful.

Change as a Constant
An important business mindset is to view change as an opportunity. There is no rational reason to automate a bad process. Therefore, look at both process and data collection as a means to improve your firm.

Finally, note that the standard approach diagram shows REPEAT as the final step. This means that change is constant and therefore, planning to review and update processes and systems based on data should be part of your firm DNA. If you are interested in learning more, check out The Business of Legal: The Data-Driven Law Practice on Amazon here.
Identifying Your Ideal Clients by Mary Juetten
The Elements of a Legal Data Strategy

The global legal community is talking a lot about the potential of software and tech tools. Others in the community are talking about the so-called “resistance” of lawyers to working with technology at all. Whether you believe there are such starkly defined “sides” on this issue, one conclusion is unavoidable: there is a sea change underway in the legal profession, indeed in the entire economy. One of the main drivers of this sea change is artificial intelligence.

But AI is not some monolithic thing. Though it is changing the world and will continue to do so, AI does not represent an unprecedented disruption of life as we know it. Technological revolutions have been common for centuries. But the spread of knowledge that occurred in the wake of the invention of the printing press, for example, did not end book production. There may have been people in dark corners claiming that the printing press meant the “death” of printing. Were they right? To an extent. Professional book printers – people writing copies of books by hand, by candlelight – probably had to change their business practices.

But on a massive scale, the world improved after book production became easier and standardized. For the most part, economies thrive from such innovation. AI represents the next great moment of endemic change.

Building a Data Strategy

Much work needs to be done to demystify a new technology when it comes along. For example, some people think AI is a giant, all-knowing robot brain that rips decision-making out of the hands of humans. This is far from the truth. AI is just software, built from data. In order for AI to draw any meaningful conclusions from data, in fact, we need a vast amount. Where do we get all that data? How do we know it’s the data we need?

To answer these questions and many others, we need a data strategy. A data strategy consists of two high-level core concepts. The first is a top-down mission statement acknowledging the value of an organization’s data. The second is a framework for developing new data-related capabilities. Think of it like a doctor first diagnosing a patient, and then recommending exercises and treatments to improve and maximize that patient’s quality of life.

A good data strategy takes an inventory of an organization’s goals, and of its resources, both human and technological. Neither people, nor AI and other tech tools, can function well in an organization where data is not managed well. However, humans and machines function superbly in tandem, with well-organized and appropriately accessible data.
Managing Legal Data

It’s easy to talk about data in the abstract, because data is somewhat abstract. At least at first. To convert data into something tangible and useful, your organization needs to consider four aspects:

- **Recognition**: Is useful data recognized as such? What is potentially being ignored?
- **Storage**: Is data being effectively stored? Is anything missing?
- **Publication**: Are there directories of all data? How are these directories structured and organized?
- **Accessibility**: Who has access to the data? How is the data used? Does everyone know how to find the data they need?

This list is unranked for a reason: a good data strategy – in legal, or in any other field – moves through these 4 stages in a cycle. Each aspect leads directly into the others, with constant improvement and iteration the ideal.

Let’s look at each of these four elements of a good legal data strategy, bearing in mind that each element is part of an ever-evolving cycle.

**Recognition**

The recognition stage of a legal data strategy represents the first place where people and technology start to integrate their respective roles. Recognizing data you have in your organization is not always easy. Oftentimes, we take a lot of basic data for granted. In documents, entity names, dates, contract lengths, and much more can be automatically recognized by trained software tools like LexPredict’s ContraxSuite. In emails, Google, Outlook, and other services keep track of metadata and content.

Data recognition doesn’t just happen at the software level, though. Legal professionals at every echelon need to work together to address what kinds of data the organization has, what kinds of data the organization needs, and whether current capture methods are getting the job done. If your organization has a lot of European clients, do you know what aspects of your agreements with those clients are subject to change under the GDPR? Do you know what your risk exposure might be in the range of that discrepancy?

**Storage**

We gather data so we can make predictions about the world. If we gather enough data, over a long enough period of time, we can be confident that our predictions will be accurate and carry significant weight. This is why data storage is so important. Recognizing good data is not enough; we also have to store our data in a meaningful, organized way. This is where technology tools like document management systems (DMS) become integral to an organization’s data strategy.

An effective legal data strategy will reliably store both structured data and unstructured data. Cloud management systems, and other forms of DMS, have made it easier than ever to store, organize, and utilize structured electronic data.
Publication

We don’t just need to recognize and store data and make it accessible, though. We need to use it. We need to produce something from it. This is another arena where humans and AI programs can work together.

“Publication” in this sense doesn’t necessarily mean that an organization is publishing and releasing company data for the whole world to see (although this often does happen in the form of SEC filings, earnings reports, press releases, white papers, etc.). Publication here refers to maintaining a directory of data within an organization. A DMS can aid in this process, but a DMS might have a default method of organizing data. This default method may or may not be the best way to communicate important data to others in your organization (e.g. some document management systems may not be able to build helpful diagrams and charts to communicate a focused message about a particular matter).

Publication is heavily focused on communication. What is our data telling us? What conclusions can we draw from an analysis of all this data? Many actionable insights come from this stage.

Accessibility

Once an organization has begun recognizing and storing data, the next stage involves normalizing the data so it is easily accessible. Unstructured data needs to be processed for specific features so that it can become structured data. Standard forms need to be discussed and agreed on in order to reduce redundancy and improve data integrity. Proper data warehousing needs to be established.

Effective accessibility means having a clear system for retrieving, analyzing, extracting, transforming, and otherwise managing data. This is largely a human-centered task; AI software can’t tell you the best way to run your organization.
Data Strategy Maturity Model
Assessing capabilities and planning improvements

Stage 1
proto-(data)-culture

Stage 2
the awakening

Stage 3
the teenage years

Stage 4
terra firma
Stage 5
golden years

Capability Continuum
reporting to analytics

Questions
Time for discussion
What is a data strategy?
Data Strategy: Defined

Statement and Framework

A data strategy is a top-down mission statement acknowledging the value of an organization’s data combined with a framework for developing data-related capabilities.

While data strategies are built on lists of principles and technologies, they address much more: strategic communication and change management, process improvement, knowledge management, and risk management, to name a few.
From Data Strategy to Wisdom

Data
- Direct record of fact, signal, symbol

Information
- Indirect record or description
- Interpretation of information

Knowledge
- Actionable inference or heuristic

Wisdom
- January is a good month to plan a ski trip to Tahoe.

Data
- Readings from a temperature sensor in Tahoe.

Information
- The average temperature in the month of December is 32.2°F.

Knowledge
- Snow is likely to accumulate in December.
When is Data Valuable?
even when it’s not

High-frequency, high-impact activities are the best use case for data
- **Systematic** understanding and treatment
- **Standardized** reporting and **statistical** treatment
- Potential for automation and prediction
When is Data Valuable?
even when it’s not

![Data Valuation Matrix](image)

**Off-diagonal** activities are often opportunities for competitive innovation and forward-thinking
- High-frequency, low-impact
  - Small efficiencies add up
- Low-frequency, high-impact
  - Model thinking
  - Positioned for future growth
When is Data Valuable?

even when it’s not

Low-frequency, low-impact activities are typically not good fits for data capabilities

However!
1. How do you justify this lack of investment without data?
2. How do you know when the frequency or impact has changed?
What makes data “big?”

(Other than the marketing department)

**Variety**
Many different “types” of data.

**Velocity**
Rate of data production or collection.

**Volume**
Total quantity of data.

**“Linked”**
Networked or relational.

🔍 **Note**
Data doesn’t have to be big to be valuable!
Context is Data

metadata is data too

Metadata is data or context about data.

For example, when and where an order was placed may be more valuable than what the order actually was.
What is a legal data strategy?
Legal Data Strategy

**Claim:** The role of legal is to facilitate business activities and transactions and to help manage and value risk.

If you accept this, then a legal data strategy should focus on developing data capabilities to improve the quality and efficiency of these tasks.
Data is often unstructured, resulting in serious barriers to accessing the potential information or knowledge within. This is especially true of legal data.

Common examples of legal data include:

- **Structured**
  - Timekeeper/billing records
  - Docket records
  - Litigation or ADR outcomes

- **Unstructured**
  - Court filings
  - Contracts
  - Deposition transcripts
Legal data strategy for corporate
justify strategies and contextualizing trends

**Justifying**
For every ten transactions or matters that the department handles:
- How many should we expect to execute or win?
- How much should we expect in liability?
- How successful are dispute resolution strategies?

**Example: ADR**
What percent of cases results in a liability >$100K?

**Contextualizing**
Are we seeing more wrongful termination disputes in Texas over the last four years? What percentage have we been losing?
Risk management for corporate

data-driven approach to answering these questions

Where does it occur?
- Work-related travel
- Workplace
- Customer location
- Reactive
- Proactive

Who Does it?
- Employees
- Customers
- Contractors

What is it?
- Workplace

Why does it occur?
- Employees
- Customers

When does it occur?
- Reactive
- Proactive
Clients are increasingly looking to their outside counsel to justify quality and fees and to handle increasing quantity. Some clients are even asking firms to bid through RFP processes or work under alternative fee arrangements (AFA).

Sophisticated firms are turning to technologies like case modeling and legal project management to meet these demands. Those that successfully implement find themselves better able to maintain margins and win business.
Process improvement for law firms

Agile

Many legal activities don’t fit into the traditional waterfall project management approach. “Agile” methodologies, originally designed for software development, are well-suited to activities like litigation or compliance. By recording strategy, duration, effort, and success across many matters, firms can leave their competition in the dust through data-driven project management.

Lean

While well-understood in business generally, Lean methodologies are relatively new to the legal sector. For non-billable activities or engagements on AFA, Lean can help identify and eliminate sources of waste, improving margins and client satisfaction. Key tenets of Lean are measurement and self-assessment, two core principles of a data strategy.
What does a data strategy enable?
Why should you build a data strategy?

five reasons to care

1. Ability to measure, monitor, and manage your resources and service providers.
2. Ability to model and improve the processes you execute.
3. Ability to allocate tasks across internal and external resources and assess cost and quality.
4. Ability to allocate tasks across risk management strategies like ADR and assess cost and quality.
5. Ability to justify and contextualize department or firm performance to the board or clients.
How do you start a data strategy?
Taking your first (small) steps

Successful data strategy implementations often begin with **small wins**. Wins often come from **well-selected early adopters**, **achievable goal-setting**, **visible use cases**, and an **open, iterative** approach.
Taking bigger first steps

In some cases, data strategy implementations align with larger organization implementation or refresh cycles.

These technology implementations present much broader, faster opportunities to enact a data strategy. However, much more effort and experience is typically required up-front than the small steps approach.
Data Strategy **Maturity** Model

assessing *capabilities* and planning *improvements*
Data Strategy Maturity Model

Maturity Assessment Dimensions

- Data Availability
- Measurement Strategy
- Reporting
- Actionability

Stage 1
Stage 2
Stage 3
Stage 4
Stage 5
Stage One

proto-(data)-culture
Data Availability

- **Recognition**: Not recognized as data
- **Accessibility**: Data is not accessible
- **Storage**: Data is not stored
- **Publication**: There is no directory or publishing of data
Measurement Strategy

- Typically not present
- Typically not refined or tested
Reporting

- Anecdotal, story-based only
- Little to no ability to validate
Actionability

- Predictive Capabilities: No data-driven predictive capabilities
- Project Management: No integration of data into project management
- Reporting and Assessment: No integration of data into department reporting and assessment
- Early Warning Capabilities: No early warning capabilities
Stage Two
the awakening
Data Availability

**Recognition**
- Increasing recognition of data

**Storage**
- Data is stored, but typically in diverse and non-normalized manners

**Accessibility**
- Data is typically difficult to access

**Publication**
- Little to no directory of publishing
- Existing data is unknown to others
Measurement Strategy

- First iterations of measurement and quality improvement begin
- Conception of metric-based thinking begins to emerge
Reporting

• First reports begin to emerge
• Reports are generally constructed in Excel
• Reports are generally not available to users outside of small requesting group
Actionability

- Predictive Capabilities: No data-driven predictive capabilities
- Project Management: No integration of data into project management
- Reporting and Assessment: No integration of data into department reporting and assessment
- Early Warning Capabilities: No early warning capabilities
Stage Three

the teenage years
Data Availability

**Recognition**
Increasing recognition of data

**Accessibility**
Data is stored, and discussions around normalization and warehousing begin

**Storage**
- Companies begin to identify and store unstructured data sources
- Data begins to be formally databased, with increasing accessibility to technical users

**Publication**
- Little to no directory of publishing
- Existing data is unknown to others
Measurement Strategy

- Master data management ideas begin to emerge around measurement standards and QC/QA
- Unstructured data approaches begin to emerge
- Existing metrics begin to crystallize
- New metrics begin to emerge at an accelerating rate
Reporting

- Begins to take on a “designated” role, with first reporting hires
- Experiments with reporting frameworks or BI applications begin
- Reports begin to be distributed outside of small groups
Actionability

- **Predictive Capabilities**: First discussions of data-driven predictive capabilities begin.
- **Project Management**: Integration of data into project management is first discussed.
- **Reporting and Assessment**: Data begins to be integrated into department reporting and assessment.
- **Early Warning Capabilities**: No early warning capabilities.
Stage Four

*terra firma*
Beginning of data directory and inventory discussions

Data is reliably recognized
- Data storage
  - Data is reliably stored
  - Data normalization and warehousing become focal point

Unstructured data begins to be processed for features
- Accessibility

Beginning of data directory and inventory discussions
- Publication
Measurement Strategy

- Master data management standards and roles emerge
- QA and QC begin to be applied to data
- Key metrics are crystallized and understood across organization
- Most “low-hanging” data has been identified, even if not yet available or measured
Reporting

- Dedicated reporting resources are available
- Reporting frameworks and BI tools receive increased investment and focus
- Reports are regularly distributed to wider audiences
Actionability

Predictive Capabilities
First experiments with data-driven predictive capabilities begin

Project Management
Integration of data into project management begins

Reporting and Assessment
Department reporting and assessment incorporates data

Early Warning Capabilities
No early warning capabilities
Stage Five
the golden years
Data Availability

**Recognition**
- Data is reliably recognized

**Accessibility**
- Data normalization and warehousing are implemented
- Data normalization and warehousing are implemented

**Storage**
- Data is reliably stored, both from structured and unstructured sources

**Publication**
- Data directory and inventory are published
Measurement Strategy

- Dedicated master data management roles and standards are present.
- QA and QC are reliably applied to data.
- Key metrics are crystallized and understood across the organization.
- “Next wave” metrics emerge, often from strategic thinking.
  - e.g., strategic recognition of “customer experience” and subsequent measurement.
Reporting

- Dedicated reporting resources are available
- Reporting frameworks and BI tools are available
- Reports integrate external data sources (e.g., industry or region baselines)
- Reports are regularly distributed to wider audiences
Actionability

Predictive Capabilities
Data-driven predictive capabilities begin to drive decision-making

Project Management
Project management is driven by data

Process Improvement
Process improvement discussions around workflows emerge

Reporting and Assessment
Department reporting and assessment is driven by data

Early Warning Capabilities
Early warning capabilities begin to develop
Historical Metrics

Tell you only **what** happened in the **past**

- Typically based on averages or aggregates, unconditioned on specific facts
- Good view of the forest health, but hard to see if individual trees/patches are doing well
- Typically show running tallies or only incorporate COMPLETED records
- Often difficult to translate into direct, actionable decisions on the ground
Historical Analytics
Can tell you why or how something happened in the past

Identify Outliers
Maybe unique, maybe areas for improvement or investigation

Not future-oriented
Can’t necessarily provide information about the future

- Ability to present more specific, focused analysis
- Ability to analyze both the forest and the trees
- Typically based on statistical methods from causal inference
Predictive Analytics
Can tell you what may happen in the future

Predictive Capability
Both YES/NO and probability/range

Predictions
Ability to predict outcome for SINGLE events

Application
Ability to reformulate as KNOWLEDGE or guidelines/policy

Machine Learning
Typically based on machine learning methods
Examples from Legal

Historical reporting

**Question:** What did we spend on settlements and legal expenses last quarter?

$1.2M

**Question:** On average, how many effort hours does staff counsel spend on the discovery phase of a non-compete dispute?

25 hours
Examples from Legal

Historical Analytics

Question: What factors drove time to close lease negotiations last quarter?

Question: What factors drove legal and compliance expenses last quarter?
Examples from Legal

Predictive Analytics

Question: Should we settle this dispute at outset?
• The counterparty is expected to accept an initial offer
• The dispute is predicted to settle for $100k, with legal expenses of $15k
• If an initial offer is not made, this dispute is expected to cost $50k in legal expenses and has a 25% chance of going to jury trial.

Question: How many staff counsel effort hours do we expect to spend on this negotiation?
• An estimate of 18 hours, with 90% confidence that the dispute will fall between 13 and 30 hours
Varying Skillsets

Historical reporting vs. historical/predictive analytics

**Historical Reporting**
- Database administrators
- SQL Developer
- Reporting developer
  - Cognos
  - BusinessObjects
  - Teradata

**Historical/Predictive Analytics**
- Data scientists and statisticians
- Data analysts
- R, Python, SAS developers
- Chief Data Officer
Baseline Comparisons
Baseline strategic decisions and resources against models

Decision-makers
- Doesn’t mean autopilot
- Suggested decisions and justification or data should be presented to users
- Users have an obligation to either:
  - Follow recommendation – agency/decision primarily owned by model
  - Reject recommendation – provide justification and explanation, incorporate into future model iterations

Analogy: Pilots
Pilots don’t get to choose any flight path; flight planning systems present optimized routes and pilots have to request an override

Analogy: CPAs
An individual accountant doesn’t get to choose what GAAP is. If they decide to calculate something in a way that doesn’t comply with the model, they have to provide a justification to the auditors.
Counterfactuals

Need to **baseline** or adjust for counterfactuals

- Example: We lost negotiations on renewal term provisions in 6/10 deals this quarter. Is this good or bad?
  - If base rate is 3/10, this is probably good.
  - If base rate is 8/10, this is probably bad.
- Example: Our average outside counsel rate rose by 1% this year. Is this good or bad?
  - If base rate is 4% increase, then this is probably good
  - If base rate is -2%, then this is probably bad.
Thank you for reading.

**LexPredict**: Legal Data Strategy Maturity

https://www.lexpredict.com

@lexpredict
About LexPredict

More About Our Company
About LexPredict

LexPredict is an enterprise consulting and technology company, specializing in the application of best-in-class processes, tools and techniques to modernize the practice of law, enhance compliance, and support legal risk management.

We focus on the goals of: (1) prediction, (2) optimization, and (3) risk management to enable holistic organizational change. We succeed in driving change by employing the right mix of people and processes, data and software, and execution and education.
Michael Bommarito

Chief Executive Officer

M.S.E.  University of Michigan
M.A.     University of Michigan

Director of Research, The Law Lab, Chicago-Kent College of Law
2015 - Present

Adjunct Professor of Law, Michigan State University College of Law
2011 – 2015

Fellow, Stanford CodeX – Center for Legal Informatics
2014 - Present

Co-Founder
Daniel Katz
Chief Strategy Officer

PhD  University of Michigan
J.D.  University of Michigan

Co-Founder

Professor of Law, Illinois Tech – Chicago Kent College of Law
2015 - Present

Faculty Affiliate, Stanford CodeX – Center for Legal Informatics
2014 - Present

Assistant Professor of Law, Michigan State University
2011 - 2015
LEADERSHIP TEAM

Michael Bommarito
Chief Executive Officer

Daniel Katz
Chief Strategy Officer

Eric Detterman
VP & Global Head of Product

Kelly Marsh
Product Owner & Consultant

Jillian Bommarito
Treasurer

Warren Agin
Director of Professional Development

LexPredict.com
Why LexPredict?

• **Industry context** (we operate exclusively within the legal industry)

• **Technical command** (we have deep expertise in the use of machine learning, natural language processing, and other advanced analytics techniques)

• **LexNLP / ContraxSuite** (we’ve built a framework to help support custom artificial intelligence projects)

• **Content** (we have a corpus of 400,000+ commercial contracts on hand to help us train and calibrate our systems and projects)

• True **interdisciplinary** perspective and expertise
Our Team Has Been Featured In the Following

A Deal Terms Study, Created Using Kira Machine Learning Contract Analysis Software

This report summarizes our findings on various relationships between materiality scrape provisions and indemnification baskets and/or deductibles. Specifically, we reviewed (1) the prevalence of and variation in materiality scrape provisions in private target M&A agreements publicly filed in 2017 and 2018 that include indemnification baskets and/or deductibles, (2) the association between double materiality scrape provisions and larger indemnification baskets (based on percentage of deal value) and (3) the association between the prevalence of double materiality scrape provisions and eligible claim thresholds.

Kira’s state-of-the-art machine learning technology automatically identifies and extracts information from contracts and comes with over 650 built-in provision models. Our team of lawyers trains Kira to find provisions by giving it example language of that provision culled from publicly filed agreements. Once Kira has sufficient examples of a provision, it can find that provision in new agreements imported into it.

Kira is not limited to publicly filed documents. With Kira, professionals can conduct their own deal point studies on the documents of their choosing—for example, a law firm’s deal documents from its document management system—to find information of interest to them. The information gathered can be shared throughout the organization and used for drafting future agreements or revising standard forms.

For a full list of the currently available M&A deal point-related provisions that Kira can identify out of the box, please see the end of this report.

Analysis of 2017 and 2018 Private Target M&A Deals

We identified a set of 89 M&A agreements from EDGAR filed between January 1 and December 31, 2017 and 116 M&A agreements from EDGAR filed between January 1 and December 31, 2018 using the following parameters: deal values between $30M and $500M, involving private targets being acquired by public companies. Transactions in which the target was in bankruptcy, reverse mergers and divisional sales were excluded. We imported these agreements into Kira and reviewed the results. Using Kira’s “Indemnification Basket/Deductible” provision model, we first narrowed our
document set to agreements containing indemnification basket and/or deductible provisions, and then used Kira to determine the prevalence of materiality scrape provisions. Kira was able to accurately identify both types of provisions, even though there was considerable variation in how they were drafted.

**Materiality Scrape Provisions**

A review of the results in our sample set indicated that, for 2017, of 73 agreements including indemnification baskets and/or deductibles, 81% also included materiality scrapes that deleted all materiality qualifications from the representations and warranties in the transaction agreement for any indemnification purpose (i.e., determining whether a breach or inaccuracy of such representations or warranties occurred, calculation of damages/losses, or both). For 2018, of 80 agreements including indemnification baskets and/or deductibles, 90% also included materiality scrapes.

<table>
<thead>
<tr>
<th>Study</th>
<th>Materiality Scrape Included</th>
<th>Not Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kira Systems 2018</td>
<td>90%</td>
<td>10%</td>
</tr>
<tr>
<td>Kira Systems 2017</td>
<td>81%</td>
<td>19%</td>
</tr>
<tr>
<td>ABA 2016-17</td>
<td>85%</td>
<td>15%</td>
</tr>
<tr>
<td>ABA 2014</td>
<td>70%</td>
<td>30%</td>
</tr>
<tr>
<td>ABA 2012</td>
<td>28%</td>
<td>72%</td>
</tr>
<tr>
<td>ABA 2010</td>
<td>49%</td>
<td>51%</td>
</tr>
<tr>
<td>ABA 2008</td>
<td>24%</td>
<td>76%</td>
</tr>
<tr>
<td>ABA 2006</td>
<td>22%</td>
<td>78%</td>
</tr>
<tr>
<td>ABA 2004</td>
<td>14%</td>
<td>86%</td>
</tr>
</tbody>
</table>

1 For example, the “Materiality Scrape” provision model achieved a 95% “recall” score, which means that out of 100 examples of that provision our staff highlighted to train Kira, it will find 95 of them. The provision model has also achieved a 91% “precision” score, meaning that out of the results that Kira finds, 91% of them are correct examples of materiality scrapes. In comparison, studies have shown that traditional word searches rarely achieve more than 70% recall or precision.
In addition to quickly assembling aggregate statistics, Kira automatically creates an easily searchable archive of the actual precedent language. A review of the results in our sample set indicated that the materiality scrape provisions are worded with significant variations. Of the agreements including materiality scrape provisions:

- 42% of the 2017 agreements and 26% of the 2018 agreements contained materiality scrape provisions that were limited solely to the calculation of damages/losses.
- 44% of the 2017 agreements and 60% of the 2018 agreements included materiality scrape provisions deleting all materiality qualifications from the representations and warranties in the transaction agreement for both the purposes of (1) determining whether a breach or inaccuracy of the representations or warranties occurred and (2) calculating the amount of any damages or losses with respect to a breach or inaccuracy of a representation or warranty (a “double” materiality scrape).
- 14% of the 2017 agreements and 10% of the 2018 agreements provided for materiality scrapes that were somewhat vaguely worded⁵ such that they could be interpreted to delete materiality qualifications from the representations and warranties in the transaction agreement only for purposes of determining the existence of a breach or inaccuracy of the representations or warranties, or for all indemnification-related purposes.
- Two agreements in 2018 included materiality scrapes deleting all materiality qualifications for all indemnification-related purposes.
- One agreement in 2018 provided for a materiality scrape deleting all materiality qualifications for purposes of determining (i) the amount of adverse consequences arising from a breach of a representation or warranty, and (ii) whether the basket had been exceeded. Similarly, another agreement in 2018 contained a conventional double materiality scrape but also included a materiality scrape deleting all materiality qualifications for purposes of determining whether the eligible claim threshold had been exceeded.

<table>
<thead>
<tr>
<th>Study</th>
<th>Materiality Scrape Limited to Calculation of Damages/Losses Only</th>
<th>Materiality Scrape Not Limited to Calculation of Damages/Losses Only²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kira Systems 2018</td>
<td>26%</td>
<td>74%</td>
</tr>
<tr>
<td>Kira Systems 2017</td>
<td>42%</td>
<td>58%</td>
</tr>
<tr>
<td>ABA 2016-17</td>
<td>43%</td>
<td>57%</td>
</tr>
<tr>
<td>ABA 2014</td>
<td>43%</td>
<td>57%</td>
</tr>
<tr>
<td>ABA 2012</td>
<td>41%</td>
<td>59%</td>
</tr>
<tr>
<td>ABA 2010</td>
<td>66%</td>
<td>34%</td>
</tr>
<tr>
<td>ABA 2008</td>
<td>32%</td>
<td>68%</td>
</tr>
<tr>
<td>ABA 2006</td>
<td>28%</td>
<td>72%</td>
</tr>
</tbody>
</table>

² This includes agreements that are silent on this issue.

⁵ For example: “The Sellers . . . shall indemnify and defend Buyer and its Affiliates . . . against, and shall hold them harmless from . . . [Losses] related to, resulting from or arising out of: (i) any breach of, or inaccuracy in, any of the representations or warranties of the Company or the Sellers contained in Article III or Article IV of this Agreement or the other Transaction Documents or any closing certificate delivered by the Company pursuant to this Agreement, in each case without giving effect to any qualifications as to materiality, Material Adverse Effect or similar qualifications contained in such representations and warranties . . . .”
We used data from our sample set to determine the association between the prevalence of double materiality scrape provisions and larger indemnification basket amounts (as a percentage of the deal value, when determinable), as compared against single materiality scrape provisions (limited solely to the calculation of damages/losses) with larger indemnification baskets. We defined larger indemnification baskets as basket amounts that were equal to or greater than 0.5% of the purchase price (“Large Basket”). Deals for which the basket amount and/or purchase price were either redacted or unclear were omitted (three deals in 2017 and two in 2018).

For 2017, of the 25 agreements with double materiality scrapes, 19 agreements, or 76%, had Large Baskets. In comparison, of the 22 agreements with single materiality scrapes, 15 agreements, or 68%, had Large Baskets. For 2018, of the 41 agreements with double materiality scrapes, 34 agreements, or 83%, had Large Baskets, and of the 19 agreements with single materiality scrapes, 12 agreements, or 63%, had Large Baskets. This suggests that transactions with double materiality scrapes tend to have larger indemnification baskets than transactions that have single materiality scrapes.

We excluded one deal that included both a double materiality scrape and a single materiality scrape.
Undoubtedly, the publicly available transaction agreements in our sample sets represent only a tiny fraction of all of the transaction agreements entered into each year that include both indemnification basket provisions and materiality scrape provisions. Using Kira, you can create a similar analysis of your firm’s own deals (and even study a specific industry or geography) with significantly greater speed, consistency and accuracy than a manual review. Kira allows you to unearth what’s standard practice for the many private transactions which are never publicly disclosed, and to easily find real precedent clause language to apply to future transactions.

We also analyzed the relationship between double materiality scrapes and eligible claim thresholds, or mini-baskets. For 2017, of the 25 agreements with double materiality scrapes, 15 agreements, or 60%, also had eligible claim thresholds. In contrast, of the 24 agreements with single materiality scrapes, only 8 agreements, or 33%, had eligible claim thresholds. For 2018, of the 43 agreements with double materiality scrapes, 21 agreements, or 49%, also had eligible claim thresholds, and of the 19 agreements with single materiality scrapes, only 5 agreements, or 26%, had eligible claim thresholds. This data indicates a strong correlation between double materiality scrapes and the inclusion of eligible claim thresholds.

Undoubtedly, the publicly available transaction agreements in our sample sets represent only a tiny fraction of all of the transaction agreements entered into each year that include both indemnification basket provisions and materiality scrape provisions. Using Kira, you can create a similar analysis of your firm’s own deals (and even study a specific industry or geography) with significantly greater speed, consistency and accuracy than a manual review. Kira allows you to unearth what’s standard practice for the many private transactions which are never publicly disclosed, and to easily find real precedent clause language to apply to future transactions.

---

5 We excluded one deal that included both a double materiality scrape and a single materiality scrape.
Clauses Kira can identify out-of-the-box in M&A Agreements

- “Knowledge” Definition
- “Losses”/“Damages” Definition
- “Material Adverse Effect” Definition
- “Material Adverse Effect” Definition — Exceptions
- “Material Adverse Effect” Definition — Exceptions — Acts of War
- “Material Adverse Effect” Definition — Exceptions — Announcement of Transaction
- “Material Adverse Effect” Definition — Exceptions — Changes in GAAP
- “Material Adverse Effect” Definition — Exceptions — Changes in Law
- “Material Adverse Effect” Definition — Exceptions — Changes in Political Conditions
- “Material Adverse Effect” Definition — Exceptions — Consummation of the Transaction
- “Material Adverse Effect” Definition — Exceptions — Disproportionate Impact
- “Material Adverse Effect” Definition — Exceptions — Economic Conditions
- “Material Adverse Effect” Definition — Exceptions — Failure to Meet Projections
- Absence of Certain Changes Representation
- Accounts Receivable Representation
- Attorney-Client Privilege Carveout
- Authorization Representation
- Bring-Down of Representations and Warranties
- Brokers Representation
- Capitalization Representation
- Compliance with Laws and Permits Representation
- Consents, Approvals, No Violations/Conflicts Representation
- Contracts Representation
- Corporate Organization and Qualification Representation
- Customers and Suppliers Representation
- Date
- Employment and Labor Representation
- Environmental Representation
- Escrow
- Exclusive Remedy
- Express Non-Reliance
- Financial Statements Representation
- Full Disclosure/No Misleading Statements Representation
- Governing Law
- Indemnification Basket/Deductible
- Indemnification Cap
- Indemnification Payment as Adjustment to Purchase Price
- Insurance Representation
- Intellectual Property Representation
- Inventory Representation
- Jury Trial Waiver
- Legal Opinions Condition
- Limitation for Punitive, Consequential or Incidental Damages
- Litigation Representation
- Materiality Scrape
- No Legal Proceedings Challenging Transaction Condition
- No Material Adverse Effect Condition
- No Undisclosed Liabilities Representation
- No-Shop
- Notice
- Owned or Leased Real Property Representation
- Parties
- Pensions and Benefits Representation
- Post-Closing Representation of Shareholders
- Purchase Price Adjustments
- Purchase Price Adjustments — Post-Closing
- Purchase Price Adjustments — Pre-Closing
- Related Parties Representation
- Sandbagging
- Sandbagging (Pro)
- Size/Purchase Price
- Specific Performance
- Subsidiaries Representation
- Survival of Purchase Agreement Provisions
- Taxes Representation
- Termination Fee
- Third Party Beneficiaries
- Third Party Claims
- Title
- Title, Sufficiency and Condition of Assets Representation
- Working Capital Adjustments

Jennifer Tsai
jennifer.tsai@kirasystems.com

Anne McNulty
anne.mcnulty@kirasystems.com

Corinne Geller
corinne.geller@kirasystems.com

+1.888.710.8454

The full text of the indemnification basket and materiality scrape provisions analyzed in this study can be downloaded on our website at https://kirasystems.com or by sending us an email at the email addresses listed on the left.

Want to try Kira on your firm’s own documents? Request a demo and arrange a proof of concept here:
http://info.kirasystems.com/schedule-a-demo
Juliet M. Moringiello  
Associate Dean for Research and Faculty Development Commonwealth  
Widener University Commonwealth Law School

Juliet Moringiello is a Professor and Associate Dean for Research and Faculty Development at Widener University Commonwealth Law School, where she regularly teaches Bankruptcy, Property, and Secured Transactions. Her scholarship addresses a variety of property issues in the law of creditors’ rights, including the relationship of state and federal law in addressing property rights and the evolution of property rights in connection with technological advances. She has published articles in the *Illinois Law Review*, the *Washington & Lee Law Review*, the *Wisconsin Law Review*, and the *Fordham Law Review*. Prof. Moringiello is the Immediate Past Chair of the Pennsylvania Bar Association Business Law Section, an elected member of the American Law Institute, and a Uniform Law Commissioner for Pennsylvania. She earned her B.S.F.S. from Georgetown University, her J.D. from Fordham University School of Law, and her LL.M. from Temple University.

Warren E. Agin  
Senior Consultant  
Director of Professional Development  
LexPredict, an Elevate Company

Warren E. Agin is a senior consultant and Director of Professional Development with LexPredict which provides uniquely qualified services in legal analytics, legal data science and training, risk management, and legal data strategy consulting. Mr. Agin also founded and currently chairs the ABA’s Legal Analytics Committee, which helps business lawyers understand how to use artificial intelligence and other analytic techniques. A practicing lawyer for almost thirty years, Mr. Agin teaches legal analytics as an adjunct professor at Boston College Law School, and is of-counsel to Swiggart & Agin, LLC, in Boston.

Mr. Agin authored the books Bankruptcy and Secured Lending in Cyberspace, 3rd Ed. (West, 2007) (the first book to discuss the use of technology in consumer bankruptcy practice), and the Bankruptcy and Intellectual Property Deskbook (ABA, 2016), and has frequently lectured on using machine learning and data in legal practice for groups such as the American Bar Association, the National Conference of Lawyers and CPAs, Massachusetts Continuing Legal Education, Thomson Publishing, and the Institute of Legal Information Theory and Techniques.

Anne McNulty  
Director of Legal Knowledge Engineering, Kira Systems

Anne manages a team of lawyers and other professionals who teach Kira, machine learning assisted contract review software, to find information in contracts and other documents. She
previously practiced M&A and securities law at Goodmans LLP in Toronto. Anne holds a J.D. from Osgoode Hall Law School where she graduated in the top 2% of her class, a B.Soc.Sc. from the University of Ottawa (magna cum laude) and a B.Com from McGill University.

Mary Juetten

Mary Juetten is the founder and CEO of Traklight, the only self-guided software platform that assesses business risk and identifies intellectual property, essentially a business legal checkup. She has dedicated my more-than-30-year career to helping businesses achieve and protect their success, specializing in transition or early stage companies to help create sustainable financial growth.

In 2015, Mary co-founded Evolve Law to accelerate the adoption of technology within the legal industry and I sold it to Above the Law in March 2018. In addition, she is a LegalShield Access Advocate. Mary holds a Bachelor of Commerce degree from McGill University and a Juris Doctorate from Arizona State University, as well as both US and Canadian public accountant certifications. Mary is currently pursing my LLM at Arizona State University’s Sandra Day O’Connor College of Law.

Mary is an international writer, speaker, and mentor, and sat on the Group Legal Services Association board and the Advisory Board of Evolve the Law. She previously represented entrepreneurs on the Board of the Crowdfunding Investment Regulatory Advocates and the Licensing Executive Society Emerging Enterprises committee.

Mary’s first book was Small Law Firms KPIs: How to Measure your Way to Greater Profits, published by Thomson Reuters in 2016, and her second, The Business of Legal: The Data-Driven Law Practice, was released in the summer of 2018, also on Amazon.

Jiaying Christine Jiang

SJD (Doctor of Juridical Science) Candidate
Emory University School of Law

Jiaying Christine Jiang is a lawyer in China and an arbitrator at Zhanjiang International Court of Arbitration. Currently, Ms. Jiang is pursuing her SJD degree at Emory Law. Her SJD research focuses on policies and regulations on blockchain technology, a comparative study of China and the United States. Her areas of interests also include comparative law, securities law, and corporate law. At Emory law, she is also serving as President of SJD Society, hosting Graduate Student Conference and conveying SJD Scholarly Writing Workshop. She holds an LL.M degree from the University of Southern California, Gould School of Law and an LL.B degree from the Shenzhen University, School of Law.