Deloitte.

Data cleaning considerations for the legal department

Data is the new oil—or so some may claim. What happens, then, if the oil turns out to be dirty? The gears of business may grind to a halt.

That's a challenge that often confronts chief legal officers (CLOs) as they try to drive change in the legal department. We've written before about the importance of managing change, but even the most robust change management program needs clean data for changes to take hold.[1] After all, there's little point to transformation if the data that legal teams work with isn't accurate, timely, readily available, or well understood.

And make no mistake, successful transformation can often benefit the legal department. Judgment and experience are necessary but likely insufficient on their own to present-day lawyering. Data can provide much needed visibility and context into your organization's legal exposure and obligations. With the right models, data can often shine a light on potential requirements, alternative outcomes, and other strategic variables. You can also use data to stress test your own assumptions or to help persuade a skeptical executive. But such insights are only as good as the data that informs them. Dirty data just won't do.

The dirt on legal data

Legal departments may lack clean, usable data for a couple of reasons. One is that, in most cases, legal isn't collecting data for itself. Instead, the legal team often relies on business systems that are meant to support other functions in the organization. Therefore, the data isn't especially useful to the CLO or the legal function in the format it comes in. Even when legal collects the data, due to differing systems, the result isn't always uniform. So are there also opportunities to make the legal department data collection process more hygienic, too?

A related issue is that legal doesn't always own the processes that could affect the company's legal posture. Contracts are a good example of this. In many organizations, in-house counsel often enters the picture only when the standard contract template is redlined, ends up in dispute, or runs afoul of regulations in a way that requires legal intervention.

Because of these conditions, clean (or useful) data can be a struggle for legal departments to acquire. Legal teams may have to manually reconcile duplicate entries and data from multiple unconnected systems every time they need to arrive at an important decision. The time and effort to do this may introduce further errors and thereby discourage team members from collecting data.
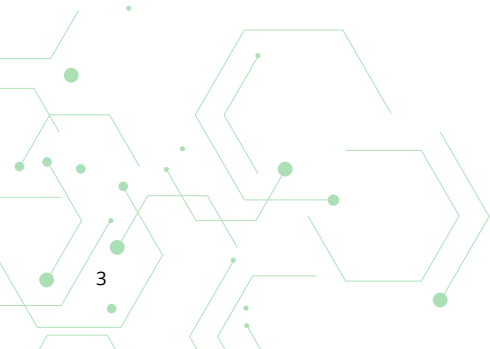
Where to tidy up

Now that we've defined the problem, let's outline a possible solution. As a starting point, consider what CLOs need to know as strategic business advisors. Although legal functions typically deal with a wide variety of data sets—regulatory, discovery, and insurance, just to name a few—three categories of business data are commonly relevant to in-house counsel and may be a good place to begin data hygiene efforts:

- Contract data, which includes customer and supplier names, expiration dates, key legal obligations, risk exposure, finance, and other regulatory compliances

- Data on current and developing intellectual property assets, including valuation, expiration dates, and competitors' filing and litigation activities

- Spending on law firms, management/ business consultants, technology, temporary help, and other budget items

In these three areas, it's helpful to have a consolidated view of the process, a way to identify status and data at each stage of the process, and the ability to track any obligations that may create risk for the organization.

Then, there's the information you'll likely want to know as head of the legal department. That includes:

- Hindsight, or operational reporting and enterprise data.

- Insight, or departmental and individual performance data.

- Foresight, or data to predict resource requirements and other needs or outcomes.

Bear in mind that not every legal process lends itself to a data-driven approach. Some are better suited than others (see sidebar). But for the ones that do, consistent, accurate, and well-defined data is paramount.

See how data and technology can help to [manage legal spend](#).

Putting data to work

Certain legal processes are particularly well-suited to a data-driven approach. These may include:

- Jurisdiction or judicial analysis.

- Benchmarking outcomes for similar matters in similar jurisdictions.

- Understanding opponents or counterparties.

- Early case assessment.

- Risk analysis (such as potential claims following an acquisition).

- Understanding and mitigating legal risk across the contracts portfolio (e.g., parent company guarantee, unlimited liability, insurance and other tax and regulatory obligations, compliance with laws).

Opportunities related to operations include:

- Enhancing outside counsel and other vendor selection to identify the right resources for a given matter.

- Controlling legal spend through alternative fee arrangements, requests for proposals, and negotiations with other organizations.

- Enhancing document review via efficiency and accuracy metrics.

- Identifying work that's ripe for new service delivery models (such as for example, moving to alternative legal service providers or applying new technology).

Organizing for data cleanup

If your organization is like most, it has a lot of data, and up to 90% of it is likely to be unstructured.[2] Prioritization is, therefore, a necessity.

Data cleanup can be prioritized along two dimensions: the value that the data has to the legal team and the cost of getting it clean (and keeping it that way). Map one dimension against the other. The data that has relatively high value to the legal organization *and* a relatively low cost of cleanup might be your top priority (figure 1).

Once you have your data cleanup priorities in order, you'll need them to become

priorities for other people as well, in order to get the job done. Given that much of the data to support legal decisions originates in other parts of the organization, consider drawing on the collaborative relationships you've formed in your role as an advisor to the business. Contract data, for instance, may exist in the following departments:

• Research and development

• Business sales

• Customer service and delivery

• Finance

• Procurement

• Inventory management

In addition, someone will need to own the data cleanup. If you outsource any of your legal processes, the service provider may be able to capture the data under their purview that you need and make it available for processing. Alternatively, you may choose to appoint a member of the legal team to take charge. If you have a cyber specialist or a records and information manager on staff, they will likely have the know how in data governance, information classification, and data mapping and inventories.

Explore what it means for CLOs to engage with other functions and business units.

Figure 1. Decision matrix for prioritizing data cleanup

High value Low cost | High value High cost

Low value Low cost | Low value High cost

Value of clean data

Cost to obtain and keep data clean
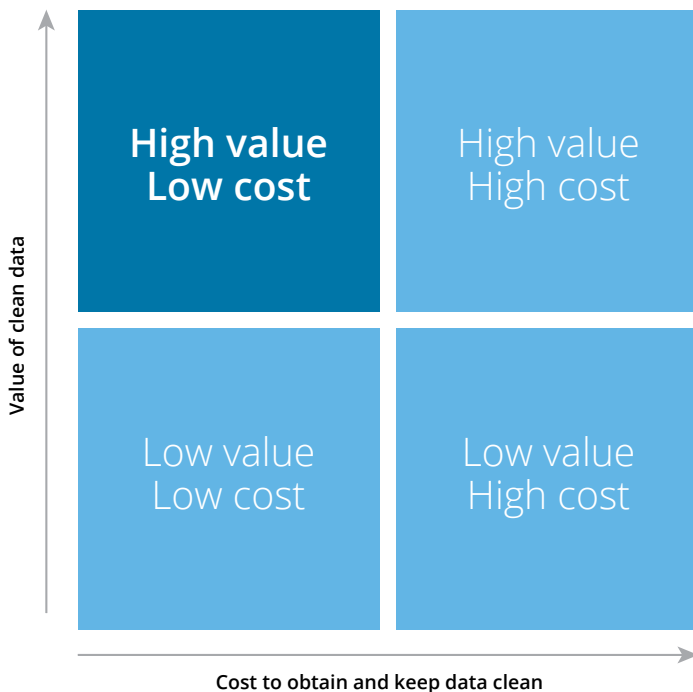
Four steps to keeping data clean

So what does it take to get data clean enough to enable sustainable change in the legal department?

Define data standards. Different legal teams may have different application requirements—think searchable contracts, e-signatures, metadata or artificial intelligence (AI)-based contract creation, and more. Common data management standards are helpful in most organizations to enable the enterprise data required for these applications to be easily deployed to legal. So keep an eye on the enterprise data dictionary, and know what your own requirements are so you can agree on the data taxonomy, data model, and other standards with a clear, useable objective in mind.

Identify your tools. Understanding the technologies on the market can help you determine how your existing data can be cleaned and integrated with future data. That includes using the defined standards, correcting metadata, removing duplications, discarding stale data, and managing data variances.

Create a repository strategy. If there are multiple systems, see if you can arrange for data to flow from one system to the other (via application programming interfaces, for example). Simplification, cost to value, time to value, and definition are likely key here. The aim is to reduce the manual intervention needed to produce a report.

Create a data curation strategy. The idea here is to simplify data entry (minimize input points), reduce the cost of data quality, and apply data cleanup practices (including AI and advanced analytics as a curation tool).

Data hygiene: A prerequisite for AI

AI is an example of a transformation opportunity that relies on data hygiene. CLOs may find that they want or need to take advantage of the productivity and risk mitigation benefits that AI tools can provide. Without good data hygiene, however, the applications and efficacy of AI can be difficult and expensive to realize. And with dirty data, the insights AI can provide—not to mention the risk exposure it can mitigate—may be elusive.

Solution: clean data **+** right tech/tools **+** right processes **+** right team

Define data standards	Identify your tools	Create a repository strategy	Create a data curation strategy
Monitor enterprise data dictionary	Understand the technologies on the market	Explore approaches to connect multiple systems	Simplify data entry
Determine data requirements for the legal team	Identify ways to clean existing data and integrate it with future data	Focus on reducing manual intervention	Reduce the cost of data quality
			Apply data cleanup practices

The data-driven decision maturity model

The following maturity model is designed to help CLOs gauge their progress on data cleanup across different legal processes. As more clean, applicable data becomes available, the more sophisticated the legal team can be in their use of data to make decisions (figure 2).

Figure 2. Data-driven decision maturity model

	Operational	Optimized	Predictive
Legal spend	Manual process Manual onboarding of suppliers on various systems Limited visibility on supplier performance Fragmented process Manual data capture and analysis	Technology-enabled onboarding Ability to track and manage suppliers and their performance Unified process Ability to generate and analyze basic data	AI-enabled automated process Technology integrated with other tools in the organization and aligned with organizational policy Ability to predict and manage the spend Predictable process Ability to structure, generate, and analyze complex data
Contracts	A single repository of all contracts Ability to manage key contract obligations at an account level Ability to meet major compliances Manual commercial management Little to no clean data	A single searchable repository with all executed contracts Defined workflow of contract approval and execution process A contracts view for renewal and expiry Use of e-signatures for document execution Basic technology used for workflow, approvals, and storage Predefined useable data	A clear view of all contracts (both buy side and sell side) across the entire organization, including: • Real-time contracting status • Team performance on each transaction • Obligation compliance • Total contract value vs. actual revenue or expenditure under each business/contract • Predictable data Ability to identify risk and mitigate it in advance
Litigation management	No usage of technologies for litigation tracking and handling Minimal litigation and reduced duration through out-of-court settlement	Use basic technologies for litigation and e-discovery process Minimal tracking on litigation Keep a balance between handling legal disputes and settling them	Thorough tracking of litigation and documentation with state-of-the-art predictive technologies Tracking of annual litigation expenditures Use of AI/machine learning (ML) technologies for e-discovery processes Strategies based on: • Prediction around litigation outcome • Litigation handling policies • Litigation expenditures and budgets • Upcoming regulatory changes and their impact

Learn more about [becoming a data-driven legal department](#).

The CLO's two hats

As we outlined in "The CLO strategist," a CLO has two hats. One is leader of the legal function. The other is business advisor, or contributor to the enterprise strategy.[3]

CLOs need to wear both hats when applying data to solve legal problems impacting the business. That makes it especially important to address processes and have the right people who can operationalize and interpret the data. Using AI to seek out patterns in your data can certainly help, but it starts with good, clean data.

Likewise, CLOs need to be mindful that data collection, no matter how well performed, is itself in service of solving a business or legal problem impacting the business. While having the right technology to collect and store your data is important, it's just a precursor to being able to make better decisions, faster, that may lead to better outcomes.

Getting fit for the future

Data has become the language of business—and, as part of the business, legal functions are expected to use data to calibrate, validate, and enhance what they do. The good news is that there's no shortage of data for CLOs to turn to. The bad news is that it isn't necessarily fit to purpose. Dirty data can be especially consequential when legal outcomes are at stake.

That's not to say data must always be perfect. Sometimes it only needs to be reasonably accurate in order to be useful. By taking a disciplined approach to data quality, CLOs can take advantage of the opportunities that a data-driven approach has to offer while laying the groundwork for the legal department's future evolution.

Endnotes

1. Deloitte, "Enable change to maximize transformation value in the legal function," 2023.

2. Tam Harbert, "Tapping the power of unstructured data," MIT Sloan School of Management, February 1, 2021.

3. Lori Lorenzo, "The CLO strategist: A new kind of legal officer for modern business," Deloitte, 2021.

Contacts

Lori Lorenzo
Chief Legal Officer Program Research and Insights Director
Managing Director | Deloitte Risk & Financial Advisory
Deloitte Transactions and Business Analytics LLP
lorilorenzo@deloitte.com

Mani Agarwal
Senior Vice President | Legal Business Services
Deloitte Tax Services India Private Limited
maniagarwal3@deloitte.com

Jack Walker
Principal | Regulatory & Legal Support | Discovery & Data Management
Deloitte Transactions and Business Analytics LLP
johnwalker@deloitte.com

Bob Taylor
Managing Director | Legal Business Services
Deloitte Tax LLP
roberttaylor7@deloitte.com

Erin Hess
Chief Legal Officer Program Research and Insights Manager
Manager | Deloitte Risk & Financial Advisory
Deloitte Transactions and Business Analytics LLP
erhess@deloitte.com

Deloitte.