

AI Today = Taming Unstructured Data

IN-HOUSE DATA LIFECYCLE MANAGEMENT SOLUTIONS ADDRESS CHALLENGE OF UNSTRUCTURED DATA.

by Alex Smith, Global Product Management Lead for iManage RAVN

The phrase “rubbish in, rubbish out” is widely used in the analytics industry, but it’s also particularly apt in the context of Artificial Intelligence (AI). Good quality data is fundamental and integral to making the “artificial” intelligent. For this reason, it’s no exaggeration to say that at this stage of the legal sector’s AI journey, the focus should be on finding and curating good data.

There is an abundance of unstructured data in law firms, but without tagging, extracting, categorising, indexing, and discovering this information, successful AI adoption is likely to be a tough ask. Companies like Uber, LinkedIn, and Airbnb have built their business systems around capturing structured data and in-house data lifecycle management solutions to address the challenge of unstructured data – and they are a success at using very well-structured data at industrial scale to provide a great customer experience.

As firms look to adopt AI and machine learning – either to unlock knowledge, build new products and services, deliver superior client experience, or otherwise provide competitive advantage – structuring and enriching data must be the number one priority. The benefits of AI cannot be properly harvested without taking this necessary first step.

What law firms need, then, is a practical approach for cleaning up and structuring data so that it forms a strong foundation for machine learning and successful AI adoption in the future.

A Words-based Data Environment

Any data cleanup project should start with the understanding that the legal world – such as law firms, corporate legal departments, alternative legal service providers, and other “lawyering” entities – is generally a words-based data environment.

This means that the “data” you’re working with largely consists of contracts, letters, memos, and other pieces of written counsel, guidance, or advice that are stored in Microsoft Word documents, emails, or PDFs. Simply put, legal data is, from a data science and information architecture point of view, mostly just a load of words. Whilst lawyers create these words in a very ordered and structured way mentally, in the construct of their

outputs to a computer, this is all lost when put into “unstructured” text formats of MS Word or PDF. This is in contrast to most other industries – like medicine or banking – whose data largely consists of actual numbers and measurements.

To be sure, structured data does exist in the legal world (think here of the hard numbers from billing systems and matter management systems) but it is vastly outnumbered by unstructured data. In facing this varied data environment, the question becomes: how best to start structuring it and cleaning it up?

Forwards and Backwards

With data cleanup, it's helpful to approach it both from a “moving forward” angle and a “looking back” angle.

What do I mean by this? The “moving forward” angle is that you start to think about the outcomes that you want to achieve, and the data you want to deliver from those outcomes – and then you build well-designed systems that capture that data. So, if you start today, in a year's time, you'll have a year's worth of structured data because you put in place a process to capture data well, which starts building a foundation for a data-driven approach.

“Looking back”, as the name suggests, is a matter of figuring how much of your existing unstructured data you want to tame. How far back do you want to go? Five years? Ten years? Thirty years? Generally speaking, you're not going to get sign-off on the time and resources required to fully and completely “own the past”.

That being said, there are certain exercises where “looking back” into the past is a necessity (LIBOR repapering comes to mind). When there's a defined business problem, this is a good opportunity to go back into the past – to find contracts and data points, and clean things up – because there is a commercial imperative to do just that.

Once you've done the data cleanup for that reason, the key is to make sure you don't get back into that mess again in the future. For most law firms, this is probably the most sensible approach: use big exercises or commercial imperatives

to go backward, and then put processes in place for moving forward.

Bringing Together Structured and Unstructured

As mentioned earlier, firms do have structured data, especially around the business of law: billing, work allocation hours, how projects are run and what timelines or milestones they hit, and so on. Most of the interesting stuff, though, is locked away in the unstructured data: the paragraphs of text in those Word documents, emails, and PDFs.

If you want to be a data-driven organisation, the question becomes how do you take unstructured data and structure it; and then, how do you marry that with the structured data you already have?

Bringing the two together is important, because if you want an answer to a question in the legal world – whether that's the business of law or practice of law side of things – that answer can't be defined by only looking at a single data source. Rich chunks of data reside in multiple systems – and the best answers can be found once structured and unstructured data are brought together.

As you get started with this effort, note that there's an overarching data architecture piece that sits over the top of the whole endeavour. Start with: what's the problem you're trying to solve? From there, ask what data would give you the answer if you could get it, and then figure out what's the best source of truth and the best system from which to draw it, whether

it's structured or unstructured. When it comes to the legal world, it pays to keep this hashtag in mind: #IAbeforeAI, which stands for information architecture before artificial intelligence.

This Isn't the Time for Amateur Hour

A word of advice when attempting any of the above data cleanup: bring in the professionals. The legal world needs professional data architects and data modellers – people who really and truly understand information architecture.

This is not the time to rely on former lawyers putting on their “innovation” hat or their “AI” hat. Legal world organisations need to hire people who do this sort of thing for a living and

have done so for years. If law firms want to move from the amateur end of innovation to the professional end of innovation, they must bring in and attract those people.

In this regard, law firms might want to take a cue from other industries. Banks are throwing massive amounts of resources behind their innovation efforts, with huge teams of professionals who are taking a deep dive into their corporate data. (In the case of banks, this is a necessity: they've had to structure their data, because there is movement afoot in the regulatory world for banking consumers to own their data. In response, banks are budgeting multiple years to clean up the data to get ready for that likely shift).

Mature innovation can also be seen in industries as varied as insurance, shipping and transportation, and even entertainment. The lesson for the legal world? Follow the lead of those industries who have taken (or are currently taking) the time to structure and clean up their data in order to better leverage AI.

No Magical Pills

A quick word here on patience. This stuff takes time. If you think about companies like LinkedIn or Airbnb or any other data-centric company of your choosing, they have taken several years to get their data in order and then make good use of it. Translation: AI doesn't happen overnight.

Unfortunately, in the last few years, the attitude around AI in the legal world has been much less patient. There's an idea that AI is some kind of magical pill that you can take, and it instantly start deriving benefits.

The truth is, there's nothing magic about AI. It's about structuring data and then finding connections and interpretation between those structured data points. That means that – as a first step – you have to think about how to turn your unstructured data into structured data. AI doesn't absolve you from doing that hard work.

With any luck, though, following the guidance and practical advice in this article will make that hard work a little bit easier, putting your organisation on the path to taming your structured data so that, ultimately, you can better reap the benefits of AI.

“This is not the time to rely on former lawyers putting on their ‘innovation’ hat or their ‘AI’ hat. Legal world organisations need to hire people who do this sort of thing for a living and have done so for years.”

“The truth is, there's nothing magic about AI. It's about structuring data and then finding connections and interpretation between those structured data points.”