

Leveraging Technology to Manage Big Data in Antitrust

February 25, 2021



[Blog](#) [Legal](#)

By Sasha Toussaint, Director of Strategic Accounts, TLS London

On 27 January, 2021, TransPerfect Legal Solutions (TLS) held the second day of the inaugural [EU/UK Competition Regulation Virtual Conference](#). [Day one focused on the changing regulatory landscape](#). Day two turned to the advent of big data and how technology is a v. tool in the arsenal of the modern competition practitioner.

[Sasha Toussaint](#), Director in TLS's London office, was joined by [Guillaume Aubron](#), Counsel at [Darrois Villey Maillot Brochier](#); [Greg Bonné](#), Senior Associate at [Latham & Watkins](#); [Bryant Isbell](#), the Managing Director of Global eDiscovery and Data Advisory at [Baker McKenzie](#) and TransPerfect's very own [Katie Perekslis](#), Vice President of E-Discovery Project Management.

Whether it's for merger control, dawn raids or cartel investigations, competition proceedings require the interrogation of significant amount of documentation production in short timelines. In the big data era, budgetary concerns as well as compliance itself often turn on the scope of the information available and timelines imposed on its production.

Clients, lawyers and regulators must fight fire with fire.

Across the board, we are seeing both clients and regulators demand the use of [technology](#). For clients, applying technology drives down the cost associated with these proceedings and limits the amount of information required for production. For regulators, it allows them to see material faster and using fewer resources. Furthermore, there is empirical evidence (in mainstream litigation versus competition work) that technology can be more consistent and accurate than human review.

In France, legal technology is not used as widely as in the US or UK. However, French lawyers see clear benefits to using technology in merger control and dawn raids. In the latter, technology needs to be deployed early, as the FCA seizes entire mail boxes before extending a fairly short time frame for lawyers to complete a privilege review ahead of their own review. Speed, therefore, is really of the essence. Unlike many other jurisdictions, there are no requirements for internal document production on the filing form itself and that really only comes into play late in proceedings. However, it is accepted that as data volumes continue to grow, technology will be a necessity to ascertain risk for these matters, regardless of the regulatory burden.

In the UK, the technology-agnostic [CMA](#) is increasingly making use of its formal information-gathering powers. Those powers are also being used at varying stages during the process – including working to establish evidence for an alternative counterfactual or a theory of harm claim. Parties similarly use technology at the “back end” regularly to formulate their strategy in response to these growing requests.

The constant across both jurisdictions (and likely others) is the use of technology to speed up the privilege review. Automating the workflow in its entirety is risky, but the technology can significantly accelerate human review.

What is considered responsive plays a large role in the documents produced in these proceedings. These decisions will have an effect on how and what data is collected and shared. Whilst the CMA retains significant control on the determination of responsiveness (described in more detail below), the [European Commission](#) takes a more balanced, collaborative approach, often engaging with lawyers and technologists.

TLS has assisted on these negotiations first hand and rather than simply paying lip service, we (together with outside counsel) have materially impacted search and responsiveness parameters. In a recent RFI, we analysed the resulting set of responsive documents based on the Commission’s proposed search terms and determined the terms were overly broad. When we presented these metrics to the Commission together with the methodology used, they were amenable to refining the search terms and data pool.

Ultimately, the regulators are amenable to changing the scope of their requests as long as the law firm and vendor work with the regulators to provide transparency and prove their methodology is logical and defensible.

If law firms can continue, as above, to guide regulators with quantitative data and defensible methodology, there will be a shift in the way data is gathered, culled and reviewed. [Deploying technology and analytics](#) is beneficial to all parties involved, particularly given its ability to better manage the never-ending increase in data volumes and sources as well as time constraints in competition proceedings.

The CMA, as compared to the [DOJ](#), [FTC](#) or the Commission, can and is asking for more extensive information earlier in the process.

As law firms and their clients lean on the benefits of legal technology, so too do the regulators. They are becoming increasingly more engaged in the process, which has [procedural and legal implications](#).

In English proceedings, [e-discovery](#) technology has been a common feature for many years but only recently in merger control. The main question focuses less on the “if” and more the “when.” The CMA is increasingly looking to gather evidence, which encourages parties to lean on the guidance and the CMA as they gather the information required. This can include internal documents and even emails and instant messaging chats. Similar to train of enquiry investigations, once they have a handle on the data they are able to issue further requests. These can be topic and custodian specific, or wider using search terms. Unsurprisingly, the CMA is actively recruiting legal technology professionals to help analyse the data that comes in to better inform its requests for information.

This is likely to cause significant delays at pre-notification resulting in a knock-on effect for the entire timetable, as was the case in [Amazon acquisition of 16% in Deliveroo](#).

Regulators, particularly the CMA, are throwing their weight around when it comes to the issue of “responsiveness.” They believe they have the power to determine what is and is not responsive through various search strategies. However, this poses an issue around personally identifiable information, documents being produced that fall out of the scope of the investigation and the general feeling that this type of information gathering request is a regulator-sponsored fishing expedition.

In general, document production in competition proceedings has become more of an iterative process and the CMA expects to be involved. As a lawyer or e-discovery professional, you can anticipate engaging with your e-discovery and forensic counterparts at the CMA every step of the way.

Whether leveraging basic technology like search terms and email threading or more [complex technology like TAR 1.0 and Continuous Active Learning \(CAL\)](#), legal teams and technologists must ensure proper processes are followed.

As technology advances, it is easy to see the benefits of using [advanced analytics, TAR 1.0 or TAR 2.0 \(Continuous Active Learning\)](#) in data-heavy matters. Regulators do not dictate what type of technology is used, but they are leaning on technologists and law firms to understand the what, how and why.



TAR 1.0 (Technology-Assisted Review) was one of the first AI-based workflows. It is an iterative process where technologists develop an initial training set of documents and a subject matter expert reviews that data before stopping review to run the algorithm that will predict the relevance of the remainder of the documents. Additional documents may be added to the training set based on the richness of the data set and a separate control set that is statistically representative of the data set is then reviewed by the same subject matter expert and used to validate the results, producing metrics such as precision, recall, depth of recall and F1 score – very much a start-stop-start-stop process.

CAL is a newer iteration of TAR and prioritises the data in real time based on the coding decisions of reviewers. As with linear review, a lawyer begins tagging documents for responsiveness. The difference is the technology is working in the background, in real time, to pull other documents that are likely similar to the previously tagged “responsive” documents until the reviewers reach a point of diminishing returns.

In TLS's experience with merger control projects where the goal is compliance to produce a large quantity of relevant documents in a short period, there is a tendency to stick with TAR 1.0 because it is not cost or time effective to put human eyes on as many documents as you would with CAL. The key in this process, however, is to make sure the subject matter expert is available to review the training and control sets to ensure consistent and accurate results because we apply their review decisions for a small population to the larger document set. Unlike a linear or CAL review, a few miscoded or inconsistently coded documents can have detrimental effects on both the quality of the review and the time it takes to reach stabilisation. Due to that limitation and a need to put human eyes on all relevant documents for fact finding, [TAR 2.0](#) still tends to be the more widely used workflow in general litigations and arbitrations given its accessibility. With either of these workflows, law firms must produce a clear methodology on actions taken, including validation process. Some of these methods include:

- Precision test – captures what percentage of documents pulled are actually relevant
- Recall test – determines if any documents were missed in the data set
- Elusion test – similar to recall and looks to see if any documents eluded the system

In TLS's experience, regulators tend to focus on achieving high recall and are less concerned with low precision (AKA overly producing documents). But precision is very important to clients, as they don't want more data to go to the regulators than necessary, particularly if the data is sensitive in nature.

All panellists agree that it is critical to produce very clear, step-by-step methodology before work has begun, both in privilege and responsiveness reviews. If the work is completed but the method is not approved, the law firm will be required to start over, which can be problematic given the tight deadlines associated with competition work.

Transparency is key in assuring the regulators that law firms have met their requirements and provided the documents required to make their decisions. As long as law firms and technologists are being defensible and reasonable about the methods, the regulators will typically approve the use of technology to work with them and with their clients.

What does the future hold?

There is a clear appetite from clients and regulators to adopt the use of legal technology in competition proceedings. Law firms and litigation support providers will continue to spearhead this movement and are constantly looking at new tools and workflows to manage big data in a defensible and efficient manner.

Law firms will leverage analytics and investigations tools where time is of the essence – the first to apply for leniency or immunity or to assess risk. Getting a handle on what their clients have early will be key.

In time, the FCA, CMA and Commission may collaborate and follow certain DOJ practices given its experience with larger data sets, deal flow and technology. And it will be in parties' best interests to work alongside them to design a process that works for everyone.

Data is inescapable. Clients, lawyers and regulators continue to play an ever-changing game of catch-up to leverage technology. The procedural and legal considerations grow and change as data volumes increase and data sources diversify, and it is imperative to lean on these tools to provide the best outcome for all parties involved.

For more information on TLS's legal technology offerings, [visit our website](#) or [get in touch](#). You can also watch the full recording of this panel [here](#).



RELATED

Legal

[de to Singapore](#)

[Advanced Managed Review: Finding the Truth](#)

s Development

By Eric Elting, TLS Director, Global Legal and Patent Business Development
12.07.2022

