Table of contents

Evolving technology and predicting the future

Mayank Sharma

ProSearch; E-discovery evangelist, enthusiast, expert

E-discovery *noun*; the process of identifying, collecting and producing electronically stored information (ESI) in response to a request for production as evidence in a lawsuit or investigation.

1. Introduction

Exercising my privilege in writing this chapter in the previous edition, I indulged my passion for movies and started by referring to Jerome Bixby's 2007 science fiction film, *The Man from Earth,* specifically as a caveat that what you were about to read was from the perspective of one person, namely yours truly. While that still stands true, this time I would like to take you back to a classic; arguably the best movie of 1957, and perhaps ever (I am of course deeply biased towards this film) – *12 Angry Men*. Ed Begley's performance as the bigoted, short-tempered, and afflicted by a nasty cough, Juror #10, has many memorable moments, but few come close to the scene where his bigotry reaches fever pitch, and as each of the other eleven jurors turns their back on him and his hate-filled rhetoric, his repeated whimpering cries of "Listen to me", lose both volume and their audience.

While my own message in this chapter is starkly more vanilla, devoid of anything remotely controversial, prejudicial, or inflammatory, I have found myself, over the years, repeating this same adage when attempting to spread awareness about the evolution of technology in the field of e-discovery. Globally, we have witnessed the fruits of technology advancement, ie, the birth of the much-hyped gig economy, the rise of AI and machine learning, the loss of people's jobs to technology, etc, *ad nauseam*. So, it isn't impertinent to suggest that it was only a matter of time before the field of e-discovery would similarly feel the effects of this phenomenon. But as we witness this evolution, might it be prudent to also look at the trends and formulate a projection of the future of this industry? I have worked as an e-discovery specialist for the last 13 years, and in this year of the pandemic, the ensuing pandemonium, and changing working landscape, I can't think of anything more pressing, critical or crucial, than this. Join me, won't you?

2.   Computers, document imaging and the internet

Before we start looking at the future, we should know where we began; namely the origin of the e-discovery industry, which is closely tied to the beginning of personal computing. While the first computer was built in 1946, the arrival of the IBM PC on 12 August 1981 was the first landmark in the journey of the industry. Kelvin McGregor-Alcorn, a personal friend, and a director in the Deloitte's Forensic Services Group, who was arguably responsible for bringing litigation support to British shores, summed up how the support industry was born:

> It was fairly obvious in the mid to late 1990s that the PC was taking over from the mini and main frame; the power of the IT Director was being dispersed and it was obvious that if your company wasn't building or selling PCs, you wouldn't be in business very long. So, I was questioning the future of the IT industry at a time when something else came along which seemed almost inevitable. To be able to review documents, quicker, better, and for less money: this seemed like the kind of proposition which was very difficult to ignore.

When I spoke to him about what he thought was the biggest disruption in the industry in its early stages, he responded with characteristic candour:

> As stupid as it sounds, one of the key moments of disruption was document imaging. Because for the first time people could look at documents on their screen, and review them on their screen and index them on their screen, rather than pulling them out of a hard copy archive, and making endless photocopies. Document imaging changed things phenomenally, and it sounds so incredibly basic, but if you took that functionality away today, and you imagine trying to manage a litigation or a regulatory intervention without document imaging, you are forced to think – how did people do it? Well, they did it, and some organisations made enormous amounts of money out of the photocopying; but it was slower, it was less efficient.

Apart from document imaging, the internet has been the most significant disruption in the field of e-discovery. With the monumental rise in the number of emails being sent daily, and the need for collaborative cross-border review, the internet has simultaneously brought challenges and opportunities to the e-discovery process. Where once a team of legal associates and paralegals would have to sit in the same room to review electronic documents on a local network, the internet made it possible for reviewers sitting on different continents to be able to access the same system remotely and work together to review information.

Production of documents to regulatory bodies in a different country at one time required the factoring in of an extra couple of days. But with the use of FTP,[1] this too became an encumbrance that could be avoided, with data being electronically transferred in a matter of hours over the internet. On the other

---

1   File Transfer Protocol, used for transferring files online.

hand, cloud storage services have brought their own challenges; for example, where individuals are able to store company data on their own private cloud storage, such storage services may come under the remit of a legal hold should the company encounter a regulatory intervention.

Webmail is another technology that tends to fall into this grey area. I remember working on a case where, among his many transgressions, an individual had allegedly used his personal Microsoft email account to transfer company intellectual property. After considerable searching, we managed to contact someone from Microsoft who was able to tell us that the only way of getting hold of this individual's email account was to directly serve a subpoena on Microsoft. Being beyond the scope of the case, this investigative avenue had to be abandoned.

3.   Processing – speed and volume

In the 13 years that I have been in the industry, I have witnessed an exponential rise in hard drive spin rates and connection and processing speeds. Almost 20 years ago, it wouldn't be surprising to take an entire working day to image a 4GB hard drive (which is smaller than a standard DVD). I remember the pain of spending a whole day imaging in 2011, but this time it was a 250GB hard drive being imaged over an eSATA connection.[2] We have now reached a stage where imaging hard drives of over 500GB can be done in a matter of hours using USB 3.0. Hard drive spin rates have gone from 4300 RPM, to 5400 RPM, to the very fast 10,000 RPM Raptor drives and the enterprise-grade 15,000 RPM SAS hard drives, and finally to solid state drives (SSDs), surpassing all (or at least most) of the above.

However, once the data has been captured, from an e-discovery standpoint it is equally important that it is processed as quickly and efficiently as possible. In the 10 or 15 years before Nuix became the industry leader in terms of data processing, a 10GB image took the better part of a week to complete. Thirty or 40GB images were considered large, and 60GB or 80GB "a couple of weeks large".

The processing turnaround times were dependent on each component of the cycle. Earlier versions of the forensic analysis tools were slow, and processing tools worked at maximum capacity of a few gigabytes every eight to 10 hours. Improvements in the underlying indexing engine technologies have been largely responsible for the drop in processing delays, while on the collection side, it was the combination of faster drives and swifter eSATA and USB 3.0 connections. This evolution in hardware and software technology has paved the way for faster, cheaper and more effective e-discovery practice.

---

2   A connection for an external storage device.

4.   Structured and unstructured

In terms of the data itself, organisations may make use of off-the-shelf solutions, bespoke software systems or a combination of both for the different areas of their business. Work product from some or all of these systems can be subject to discovery.

'Structured' data refers to clearly defined data types such as relational databases or financial systems data, while 'unstructured' data is normally everything else, for example, emails, electronic documents and Sharepoint or network shares. The majority of data an organisation generates, as you might expect, is unstructured; according to industry analyst Seth Grimes, "eighty percent of business-relevant information originates in unstructured form, primarily text". Suffice to say, important case-relevant information could just as easily be part of the 20% structured portion (perhaps buried in transactional form in a large sales ledger) and therefore each of these two categories comes with its own set of challenges.

Thus, a rigorous scoping exercise is crucial at the start of a case, where the e-discovery professionals and the client and counsel define and refine the data universe for a given case in collaboration. Questions about archiving and backup policy, where an individual is allowed to store data, remote access, need to be asked at this stage to get as clear an idea as possible about the proportion of structured and unstructured data that may need to be collected. In the case of structured data, this is also the time when conversations around the method of collection, whether in the form of large database exports in text format, or archive files, should take place. On the unstructured side, it is equally important to have a clear idea of what level of access users have, where they store data, and the level of permissions they have in terms of transferring data outside or using third-party applications to create work product.

This is an extract from the chapter 'Evolving technology and predicting the future' by Mayank Sharma in International E-Discovery: A Global Handbook of Law and Technology, Second Edition*, published by Globe Law and Business.*