

Big Data: What is Big Data? Where is it found? What are the challenges and opportunities it represents?

By Anton Hughes

Big Data is becoming increasingly prevalent in our society. Because of this, it is important for barristers to know what it is, and how to use it.

What is Big Data?

A definition of Big Data does not depend on an assessment of physical size. There is no need to measure the size of a dataset to see how much storage space it takes up, before deciding that 1.6Gb is suitably big. Instead, the usual definition of big data relies on it meeting one of the following criteria (known as the 'three Vs'):

- a. **Volume:** any data set so large it presents challenges in its storage, or processing;
- b. **Velocity:** data is generated at a rapid pace, and there is a need to respond to it in real time; or
- c. **Variety:** data is stored in a variety of different formats.

Where is it found?

As a general proposition, big data is stored in databases. Beyond databases, an organisation's Big Data may also be available in Word documents, email servers, mailing lists, social media profiles, and a business' website.

Challenges with Big Data

Since barristers rely on client or personal data as a source of value, we must also confront potential security issues. Data can be compromised or stolen due to such things as:

- a. inadequate security controls
- b. malicious insiders;
- c. external threats; and
- d. weak system security configurations.

Understanding this potential risk requires the assistance of an expert information technology professional, and preferably one experienced in security assessment.

A related issue arises here, in that any breach of data security might give rise to a breach of client privacy. Where you have any sort establishment in the EU, or offer goods or services, or monitor the behaviour of individuals in the European Union, the *General Data Protection Regulation* (GDPR) contains data protection requirements that apply from 25 May 2018. Any Australian business has to comply with the requirements of the *Privacy Act 1988* (Cth), which are similar but not identical.

Opportunities with Big Data

Barristers are in the business of dealing with Big Data, as anyone who has received a 5 volume brief can attest. So our individual capabilities to analyse and assess large amounts of data and extract valuable insights is a reputational advantage we might all exploit.

However, there is a real opportunity for individual barristers to distinguish themselves by becoming better at using existing tools that are available to make the information extraction process more efficient.

At the most basic level this might just involve knowing how to properly use optical character recognition (OCR) to extract text from scanned documents, so you can search and find text in them. Or it might involve being able to use Excel's pivot table feature, or even using Excel to present data in chart form. Web applications like Multiplottr can plot a list of addresses on a Google map.

At a more advanced level, tools like Tabula can be used to extract tabular data from PDFs. BarNet Jade can read your submissions and create a list of authorities from it. Orange can extract key phrases from text, like places, people, locations or events from documents.

*This post was originally published on [*Counsel of Tech Perfection*](#)