



## STRUCTURED DATA: Issues and best practices

Recent amendments to the Rules of Civil Procedure mean issues like spoliation, sanctions, and adverse impacts are focus areas for many attorneys, providers, and clients.

David Turner, a Senior Managing Director in our Data & Analytics practice, discusses the issues that are often overlooked, and describes the technological best practices regarding preservation and proportionality, in particular the challenges associated with client's structured data.

### Why is structured data important?

"Electronically stored information" (ESI) usually refers to unstructured data such as emails, text messaging, electronic document files, and social media messages. Yet this is just the tip of the iceberg. Around 70% of a company's information is maintained in structured forms such as records in a relational database, or in semi-structured hybrid formats such as in Salesforce.

This data is critical to understanding all aspects of an investigation. For example, when discussing whether Trader A intended to manipulate commodity prices, it will be necessary to analyse potentially hundreds of millions of transactions in order to answer questions such as "Did their trades have the effect of manipulating prices, and if so what was the price effect of this manipulation?" If the issue is whether Broker B was trying to front-run customer trades, analysis of structured data could address the question, "Were their trades executed before customer trades?"

### Getting ahead of the litigation wave through best-practice data preservation

There is a lot that can be done to get ahead in preservation before getting to the point in litigation where you are engaging counsel and hiring a third party service provider. In particular, strong information governance makes preservation much more efficient and successful. Best practices would include:

- **Identify** – know what data there is using data maps
- **Transfer** and aggregate the data (so all information is available in one place if a case hits)
- Create a **directory** to help review the location of data (for example, if it is with Counsel)
- Determine the **relevant population**
- Assess **redundancy** needs, considering defensible deletion for duplicated data to reduce storage costs and risks

To take a couple of the aspects in more detail, if we consider redundancy, the disposal of data has multiple benefits. Although it is necessary to ensure that important data is preserved, keeping 30 copies of it has no benefit. Disposing duplicated data can reduce both, costs and cyber security risks.

Adopting information governance best practices across the board will improve this process, as well as reducing risk and cost and improving data security.

### Structured data and preservation

The best practices discussed above apply to both unstructured and structured data, although structured data requires special handling. For example, it is necessary to:

**Identify all the sources of potentially relevant data.** This applies especially to legacy data. If a system was migrated in 2007, did all the required historic data come with it? If not, it may be necessary to go to an offline archive.

**Preserve dynamic data immediately assuming a litigation hold.** It may be necessary to suspend routine data purges, which can require some system reprogramming. Backup procedures can be modified to ensure required information is kept longer to meet preservation needs. There is also the option of creating copies of relevant data files. Whatever procedures are adopted must be adhered to, and must be captured correctly so that they can be described.

**Preserve reporting options.** A database can't be simply opened up and reviewed as if it was an email. Therefore, reports should be preserved and should provide a snapshot of the data at the time the report was run, together with an indication of the data shown to those receiving the reports.

**Determine parameters for gathering responsive data.** This can be complex because databases tend to contain codes in place of recognisable keywords. To find everything that satisfies a given criterion, it may be necessary to write and run scripts. During the preservation period, the location of the data dictionary and entity relationship diagrams should be ascertained for every database that may contain responsive information. Preparing representative samples from databases can preempt potential problems.

### Structured data and proportionality

Proportionality – ensuring you only produce the data that you need to – helps manage costs and risks. It can cost around \$18,000 to review a gigabyte of data. Even though storage costs are reducing, storing a terabyte of data for a year can still cost around \$3,200, so those costs can quickly mount up too.

Predictive coding – the use of keyword search, filtering and sampling to automate portions of the review process – is a great way to do more for less when it comes to reviewing unstructured data, and is rightly being increasingly accepted. However, predictive coding is not usually applicable for structured data, which requires a deeper understanding of the universe of information.

Yet structured data is associated with proportionality issues of its own. It's necessary to find ways to filter the data without the ability to use keyword or concept searches, as well as to produce the data in a format that can be reviewed by attorneys.

Fortunately, technology exists to help with these issues. Advanced analytics, data mining, and visualisation tools, in particular, can effectively harness value from structured data. For example, it's possible to provide a customised structured data redaction tool that enables an attorney to review general ledger data in much the same way as a document, maintain multiple versions of privilege and PII redactions, and produce it in 'near native' format. Visualisation technology is helpful in explaining this approach to clients, adversaries and judges: for example, showing where "relevant" data comes from and why a given approach to production is defensible.

### Best practices for structured data production

**Know your systems.** When dealing with SAP, for example, take advantage of viewer extraction tools that don't require users to deal with large numbers of tables.

**Look for a "single source of truth".** All necessary information may exist already in a data lake or repository with feeds from several operational systems. Identifying such sources is a massive time-saver.

**Think about production formats.** What will data look like if it's produced for the other side? Working backwards from how it should look may reveal the best way of extracting and collecting it from the source.

**Get close to the IT team.** During information governance and the discovery process, particularly of structured information, it's essential to work closely and proactively with the IT team. This team needs to be aware of the process, of what is expected of it, and of the potential consequences of failure (such as spoliation, sanctions and adverse inferences).



Nick Hourigan  
Senior Managing Director  
London, UK  
T: +44 20 3727 1343  
nick.hourigan@fticonsulting.com

Paul Henninger  
Senior Managing Director  
London, UK  
T: +44 20 3727 1845  
paul.henninger@fticonsulting.com

EXPERTS WITH IMPACT™

#### About FTI Consulting

FTI Consulting is an independent global business advisory firm dedicated to helping organisations manage change, mitigate risk and resolve disputes: financial, legal, operational, political & regulatory, reputational and transactional. FTI Consulting professionals, located in all major business centres throughout the world, work closely with clients to anticipate, illuminate and overcome complex business challenges and opportunities.

The views expressed in this article are those of the author(s) and not necessarily the views of FTI Consulting, its management, its subsidiaries, its affiliates, or its other professionals.

[www.fticonsulting.com](http://www.fticonsulting.com)

©2017 FTI Consulting, Inc. All rights reserved.