Approaching Different Data-Types in UK eDisclosure

Co-authored by Alejandro Gomez-Igbo, a Director in FRA's Data Governance & Forensics team and Fiona Campbell, a Senior Associate in Fieldfisher's Dispute Resolution Practice.

Fiona and Alejandro have considered some holistic factors for practitioners working with general data-types in the context of UK eDisclosure matters, with particular attention paid to ever-increasing structured data types.

UK Disclosure – What are the Rules?

Disclosure is a part of the litigation process whereby parties to proceedings are required to produce to the other party/parties, documents within their possession or control that are relevant to the issues in dispute.

Depending on the jurisdiction of a matter, disclosure in the UK Courts will be governed by Civil Procedure Rules 31, Practice Direction (PD) 31A and 31B or Practice Direction 57AD (PD57AD).

Potentially relevant documents within a party's possession or control must be preserved as soon as it becomes apparent that litigation may ensue.

Documents fall into three disclosure data-types:

- Structured data
- Semi-structured data
- Unstructured data

As each of these data-types differ in nature, so too will the manner in which they should be collected and reviewed.

Unstructured data is typically what eDisclosure practitioners see on a daily basis, including word documents and emails. Structured data and semi-structured data do not follow exactly the same steps as unstructured and require a different approach to ensure the integrity of the data is maintained. In addition, depending on the context and objective for using the structured data collected, it needs to be either reviewable in a user-friendly format or presented in a visually concise and informative manner when subject to disclosure.

What are the differences in eDisclosure Data-Types?

What is structured data?	What is unstructured data?	What is semi-structured data?
Structured data refers to highly-organised, formatted data typically residing in complex applications or databases, such as ERP systems, proprietary systems, legacy databases, financial & accounting systems, and customer relationship management systems. Structured data can be displayed in rows, spreadsheets, relational databases, and includes highly-organised data that is easy to analyse.	Unstructured data can be textual or non-textual and includes typical standalone data formats we work with in our everyday lives, such as e-mail messages, word documents, excel files, presentations, instant messages, video and audio files. Unstructured data is commonly sought in disclosure exercises.	Semi-structured data is a middle-ground between structured and unstructured data. It includes things like social media posts where the content of the post is unstructured, but the associated date and time information, user interaction (such as likes and comments) are structured.

eDisclosure tools have been designed with unstructured data in mind and innovative techniques including deNISTing, de-duplication, email threading and optical character recognition to name but a few, allow eDisclosure professionals to quickly eliminate irrelevant unstructured data to provide a document review corpus that is more likely to be relevant than not. Furthermore, automation tools and machine learning can be used to expedite reviews of unstructured data. Structured data oftentimes does not follow the same process and requires a different skillset.

Structured and semi-structured data records, such as financial transactions, often contain data relevant to the issues in proceedings. Relevant structured data can be identified by forensic experts using advanced analytics, including the use of machine learning where appropriate. The AI algorithms used in addressing structured data can show hidden patterns and systematic behaviours within a structured dataset that may not otherwise be observed by traditional methods. For financial litigation in particular, there is a requirement to combine structured data with unstructured data to identify not only what happened in a transaction, but the decisions and discussions behind the transaction. Organisations should know and understand what structured data resides within their environment. In order to do this, a comprehensive data scoping exercise should be actioned. At this stage, it is beneficial to engage the services of an expert team to properly assist with general data scoping to ensure that all potential structured data-sources are considered, particularly if third party providers are involved and/or if there is a cross-border element to the geographical location of the structured data.

Complying with UK Disclosure Rules

In-scope relevant documents may be held by employees or in central repositories such as archives, or on servers and back-up systems. As such, the duty to preserve documents extends to third parties, including employees, former employees or agents.

Clients are required to confirm to the court and to the other party/parties involved in a dispute that the necessary preservation steps have been conducted in line with court rules. To comply with court rules, a robust plan considering an organisation's full data landscape should be put in place at the earliest opportunity of any eDisclosure exercise - this plan may necessitate the assistance and involvement of an organisation's subject matter experts, including IT teams and archivists, as well as legal eDisclosure experts and external forensic eDisclosure professionals. Once preserved documents have been collected, it is best practice for clients to also confirm to their legal representative that all known in-scope data-sources have been collected – this can be conducted with the assistance of an eDisclosure advisor.

Failure to properly preserve data or where data is inadvertently destroyed, is likely to expose a party to court sanctions.

eDisclosure Cost Factoring

eDisclosure rules require parties to estimate the associated costs of their eDisclosure exercise, including the costs of technology and eDisclosure advisor assistance. PD57AD in particular, requires parties to conduct 'reasonable and proportionate' searches for relevant data – what is reasonable and proportionate is gauged by the nature and complexity of the case, the number of documents potentially involved and associated costs.

The role of eDisclosure advisors in correctly assessing and collecting data sources is important, alongside any methodologies they can employ to accurately assess data content for relevant information and simultaneously omit irrelevant data. Whilst upfront eDisclosure costs may seem high, it is important that this process is conducted in a defensible and effective manner to avoid court-imposed costs for improper disclosure at a later date.

Addressing Different eDisclosure Data-Types

The eDisclosure life-cycle for all of these data-types will generally follow the same format:

1. Data Identification and Landscaping
2. Collection
3. Processing
4. Early Case Assessment
5. Search and Review / Analysis
6. Production

Key Points for Data Collection

While unstructured data formats typically rely on industry-standard tools and processes to preserve and extract data, structured data platforms usually require input and dependency from a system administrator of the data repository. This leads to a dependency on the organisation subject to the dispute to execute (or at least form a part of) the data extraction exercise. Oftentimes, the metadata of a structured dataset may not undergo the same scrutiny in a disclosure exercise to that of unstructured data. However, ensuring there is no spoilation of data during extraction, transfer and analysis of data is viewed as best practice and any data forming part of a disclosure exercise should follow the "ACPO Principles of Digital Based Evidence". This can be achieved by ensuring that eDisclosure experts are involved as early as possible during the data scoping exercise, as a system administrator may not have the appropriate experience and expertise to extract the relevant data in a forensically defensible manner. For example using cryptographic hash functions when transferring files to highlight data integrity, tampering or corruption issues.

In terms of scope, systems can be wide ranging and complex within an organisation, with simple reporting tools such as financial systems, HR systems, customer data etc. all sitting within disparate, unconnected systems. For that reason, it is of high importance to conduct data landscaping in order to fully understand the layout of the organisation's environment so that the scope can be clearly defined. Once this is determined and extractions commence, casting a wide net can prove to be beneficial should the scope of analysis change throughout proceedings. For instances where entire databases are preserved, data dictionaries and entity relationship diagrams are also necessary to provide clarity on what the multiple tables and fields in database work and relate to each other. While casting a wide net is typically the preferred and most defensible method for preserving data. A more targeted approach can also be taken at the outset of the preservation exercise to gain an early understanding of the data landscape and produce initial observations while the full dataset is processed. With that in mind, it is important to understand the limitations of how the targeted dataset was produced so that any insights derived from it are not misinterpreted.

Another consideration to be made surrounds legacy data and data migrations; matters that cover a large time period may require data from legacy systems which may no longer be live. In this case, it is imperative to understand the use of the legacy systems, if they were migrated and potential ways of restoring the data if it only exists on a legacy system - this is a key consideration to understand early on in the disclosure process, as restoration of legacy systems when required can be a time-consuming and a potentially costly exercise.

Importance of Getting in Early and Getting it Right

While some of these tasks may seem time-consuming. The importance of following the right steps is more crucial than ever. The data scoping and collection phases are becoming increasingly more fragile and the knock-on effects of a mishandled scoping and collection can go unnoticed until much later.

Exponential growth in the volume and type of electronic data within organisations means that if disclosure requests are not addressed properly and early when disclosure becomes apparent, then risks and challenges will be posed further along in the proceedings. Tackling the process in an ineffective manner may have adverse cost implications, with courts and regulatory bodies imposing penalties for bad practices, coupled with that the reputational damage for conducting disclosure improperly may have a greater impact to an organisation.

Contact us

Alejandro Gomez-Igbo
Director, London
Forensic Risk Alliance
agomezigbo@forensicrisk.com

forensicrisk.com

Fiona Campbell
Senior Associate, London
Fieldfisher
fiona.campbell@fieldfisher.com

fieldfisher.com