# REMEMBER THE "GOOD OLD DAYS" OF EDISCOVERY?

**Dennis Kiker**   Granite Legal Systems

Remember when the new eDiscovery amendments to the Federal Rules of Civil Procedure were actually new? We are now more than eight years beyond that fateful day when eDiscovery became part of the regular legal parlance. Barely a decade ago, many cases (perhaps even most) still focused on paper records, and electronically stored information (ESI) was the new frontier. Today, it is a rare case when parties do not seek and produce email, word processing files, spreadsheets and presentations.

However, an increasing volume of potentially relevant information is not contained in those routinely produced documents but in enterprise information systems, structured databases, and legacy applications. These data repositories are more difficult to access and collect from, and more difficult to transform into a format suitable for attorney review and production in litigation. Absent a disciplined, methodical approach, discovery from these "complex data sources" can cost companies an inordinate amount of time, money and resources. This article outlines a defensible process for discovery from new sources of ESI. Although you can apply the same strategies to virtually any new data source, such as mobile devices and web-based applications, we will focus here on structured data.

First, at the risk of going geek for a moment, let's make sure we understand just what "structured data" is. IT professionals refer to office documents and email, respectively, as "unstructured" and "semi-structured" data, referring to the way in which the data are stored in the computer system.[1] Unstructured data are not organized according to any predefined schema. Every document is a self-contained entity and the only organizational structure around its storage is generally that created by the user (e.g., folder structures). Email is somewhat similar, though the data include tags and other markers to facilitate organization and retrieval. "Structured data," on the other hand, is, well, structured. In most cases, we are talking about a database or other enterprise system in which data are organized according to a defined schema. Examples of structured data include customer complaint databases, manufacturing quality systems, and accounting systems.

Why is structured data important? Because it makes up a substantial percentage of business information. According to a recent study by Tata Consulting Services, companies store anywhere from 33-67% of their data in structured formats.[2] Imagine ignoring up to 67% of the information available to you when investigating a matter. Obviously, not all structured data will be relevant to your case, but some of it may well be important to the defense or prosecution of your matter.

Moreover, just as parties were battling over the necessity of producing unstructured data eight years ago, databases are coming into play, sometimes with unfortunate results. For example, in *Advanced Tactical Ordnance Sys. LLC v. Real Action Paintball, Inc.*, No. 1:12-CV-296 (N.D. Ind. Feb. 25, 2013), the plaintiff sought production of a database that, according to the defendant, "operate[d] all aspects of RAP4's business," and contained the defendant's "customer list, including all information about its customers; all orders for its products; its marketing programs and newsletters; its customer service/support ticket procedures; its implementation of customer vouchers; frequently asked questions; and information about its dealers and special programs made available to them."

Although the court acknowledged that the plaintiff's request "appear[ed] facially intrusive," it held that the benefits of allowing access to the database outweighed the burden of producing it. *See also Chen-Oster v. Goldman, Sachs & Co., et al.*, 10 Civ. 6950 (AT) (JCF) (June 18, 2013) (resolving privilege issues over previously ordered production of data from two databases, including the defendant's "comprehensive repository for human resources information" and the database that contained "the results of the annual year-end compensation review process").

Of course, courts will not always order the production of entire databases or even

raw data extracted from databases simply because a party asks for it.[3] Nevertheless, considering the volume of information in enterprise databases and the likelihood that some of that information is probably relevant to your case and perhaps even advantageous to your client, it is important to have a defensible methodology for collecting and producing such information. The following methodology, applicable not only to structured databases but other data sources as well, consists of three major steps: assessing the source, defining and executing the collection methodology, and developing a data transformation strategy to enable attorney review.

### ASSESSING THE SOURCE

Too often, attorneys and their clients move directly to collection from structured data sources without taking the time to critically assess the repository. As a result, the data collected and its format may not meet the needs of the litigation. This is because most information systems are not designed with litigation in mind. Thus, the information accessible through standard user interfaces might not include information that is important to the case. For example, in a complaint handling system, standard reports may not include important metadata, such as information about the dates on which particular entries were changed. Without first assessing the data source, it is impossible to know what you don't know about it.

Key details to document include:
- The application software and database type;
- The types of data maintained in the application or repository, including fielded information (data entered directly into the application by a user), metadata (system generated data), and attachments (documents uploaded into the application and associated with a "record" or unique set of fielded data);
- The volume of data by data type;
- The source(s) of data contained in the system;
- The data management and archive policies;
- Access and security issues;

- Reporting functionality;
- Data extraction capabilities.

The outcome of this exercise should be a detailed system analysis profile that can be reused in subsequent litigation.

### DEFINING AND EXECUTING THE COLLECTION METHODOLOGY

With a complete profile of the data repository, you can then identify potentially relevant information and determine the most efficient method of acquiring that information. On occasion, this involves using standard user interfaces available with the application, but, in most cases, the more efficient course will be to extract the raw data from the system and format it externally. There are two primary reasons for this: speed and flexibility.

When the volume of data is large, accessing it through the user interface (the "front end") is often too time-consuming and burdensome. Many applications are simply not designed for large-scale extraction of data from the front end, which is generally designed for single transactions (e.g., looking up a single complaint record and generating a report). Generating large volumes of data, therefore, often will require a great deal of human interaction, which is not only disruptive and time-consuming, but generally more costly.

In contrast, it is often possible to programmatically extract data from the back end, reducing both the time and cost of obtaining the data. Moreover, extraction of raw data from the back end provides more flexibility in the production format. For example, where a database contains information that needs to be redacted prior to production, producing reports from the front end generally means that the resulting documents must be manually redacted one by one. On the other hand, it is possible to structure and programmatically redact raw data extracted from the back end, resulting in improved accuracy and efficiency.

### DATA TRANSFORMATION

Much of the data in a structured database is simply not susceptible to attorney review in its native state. By design, databases

associate various pieces of information as required for various purposes. For example, in an order fulfillment system, there may be information identifying customers, the products that they have ordered, and the orders that have been shipped. Rather than include all of the information known about the customer with every record of an order or shipment, the database will store the customer information once and then associate that information with order and shipment records using a customer number. Thus, shipment data alone will not enable you to identify the customers.

As a result, it is necessary to transform the data into something that is readily reviewed by attorneys for relevance and other issues. This can be accomplished in a variety of ways, the selection of which depends on the nature of the data and the objectives of the review.

### CONCLUSION

The Federal Rules on eDiscovery are no longer new, but that does not mean there is nothing new about eDiscovery. One of a number of new challenges in this area is structured data. The next time you are presented with a new matter, in addition to the routine eDiscovery that you'll undoubtedly think about, also consider non-standard information, including structured data. It is increasingly likely that there is some structured data relevant, and quite possibly advantageous, to your case, but it requires a disciplined approach to identify, preserve and collect that data without incurring unreasonable cost.

*Dennis Kiker is a Consultant at Granite Legal Systems in Houston, Texas, specializing in eDiscovery consulting and technology. He is a Martindale-Hubbell AV-rated attorney and legal consultant working with law firms and their clients to facilitate and improve discovery response, including the development and endorsement of internal and external processes, and the identification, preservation, collection and production of information in active litigation. He is a member of the State Bar of Arizona and the Virginia State Bar. He can be reached at dkiker@granitelegal.com or 713-652-0881.*

---

1　See, e.g., "Semi-structured data," Wikipedia (2013), *available at* http://en.wikipedia.org/wiki/Semi-structured_data (last visited on 1/14/2014); "Unstructured data," Wikipedia (2013), *available at* http://en.wikipedia.org/wiki/Unstructured_data (last visited on 1/14/2014).

2　"Big Data: A TCS 2013 Global Trend Study" (2013), *available at* http://sites.tcs.com/big-data-study/tag/unstructured-data/ (last visited on 1/14/2014).

3　*See, e.g., 150 Nassau Assoc. LLC v. RC Dolner LLC*, 2011 NY Slip Op 30337 (N.Y. Sup. Ct. Feb. 14, 2011) (*available at* http://www.courts.state.ny.us/Reporter/pdfs/2011/2011_30337.pdf) (denying production of database in "raw" or native format where the same information had been produced in a reasonably usable format).