

# Data Dumps: The Bane of E-Discovery

By Sharon D. Nelson, Esq. and John W. Simek

© 2011 Sensei Enterprises, Inc.

Everyone **knows** you're not supposed to do a data dump in e-discovery. But oh boy, is there a temptation to drown the other side in a case with an avalanche of useless data. Too often, law firms and their clients succumb to this temptation.

In *SEC v. Collins & Aikman Corp.* (S.D.N.Y. 2009), the SEC dumped 1.7 million records (10.6 million pages) on the defendant saying that the defendant could search them for the relevant evidence and asserting that it didn't maintain a document collection relating specifically to the subjects addressed. As the court correctly noted, Rule 34 of the Federal Rules of Civil Procedure prohibit, "simply dumping large quantities of unrequested materials onto the discovering party along with the items actually sought." The court also found that asking the defendant to do the plaintiff's work involving a huge outlay of time and money constituted "undue hardship by any definition." The Court ordered the SEC to perform its e-discovery duties in accordance with the rules.

More recently, in *Felman Prod. V. Indus. Risk Insurers* (S.D. W.Va. July 23, 2010), Plaintiffs admitted that nearly 30% of their production was irrelevant. As the court noted without humor, the production included "car and camera manuals, personal photographs, and other plainly irrelevant documents, including offensive materials." So the judge took the Plaintiffs to the woodshed. Having produced thousands of attorney-client documents inadvertently in what the judge called a "ridiculous" production, he found that Plaintiff's review and production methodology was not reasonable and that the attorney-client privilege had therefore been waived. As an added bonus for Defendants, the production was so sloppy that there were a couple of real gold nuggets in the now non-privileged attorney-client e-mails. We are quick to note that a software glitch may have caused some of the problems in this case, though proper review should have caught it – and some experts have challenged the judge's math, but clearly this was not a case in which all the rules were followed. The judge's irritation with over-production is consistent with the mood we have seen on the bench.

It is fairly common to hear complaints about federal government data dumps. In *U.S. v. Stevens* (D.C.C., Defendant's Motion to Compel Discovery, Sep. 2, 2008), the Defendant complained that the government had produced thousands of documents in an unusable format that "appeared to be an undifferentiated mass, with no discernible beginning or end of any given document." As courts tackle this issue, it is becoming clear that litigants must label the documents produced in response to requested subject areas. Data must be organized, searchable and indexed. Obfuscation is not acceptable production.

In criminal law, attorneys frequently report to us that the prosecution will do a data dump on defense counsel, effectively burying any exculpatory information in a sea of data. Several courts have noted that a deliberate data dump, done for the purpose of avoiding adherence to *Brady* obligations, would not be permitted. This is an area ripe for clarification, as many defense lawyers have reported that prosecutors are "opening their files" to the defense rather than specifically providing exculpatory information.

Most of the buzz is in the civil world where it is widely alleged that large law firms use data dumps to overwhelm small law firms. And perhaps so. But the real question is: How do we prevent this abuse of the e-discovery process?

For one thing, the Meet and Confer happens way too late. That's why everyone is turning to "early case assessment" which has become a buzz phrase. The minute you know you're involved in litigation (or likely to be), you're under a litigation hold. Now you have to decide what to preserve, preservation being broader than production. Already you need to do three things:

1. Retain an e-evidence expert (we wish they would, but often they hire an expert much later, when everything is now an emergency and cost-saving advice is coming very late in the game, after way too much has been spent already).
2. Talk to your opponent and start getting consensus about the scope of preservation on both sides (which means the exchange of a lot of info)
3. Within the litigation hold team, begin early case assessment.

Who are your key players and what sources of data do they have (workstations, laptops, home machines, smartphones, voicemail, flash drives, etc?) What other data may be relevant? Do third parties hold data? What's the likely volume of data that will be preserved? How can it be winnowed down?

It is never too early to talk to the other side about the format of data to be produced or to begin talking about search methodologies, although that often occurs at the Meet and Confer. From early case assessment through the Meet and Confer, the ways to reduce data volume should be at the forefront.

Native format is both cheaper and the "best evidence." You often need searchable PDF (or TIFF with load files) in order to redact/Bates stamp. Requesting a "mixed" production (primarily native) is perfectly acceptable. You can agree on rolling productions if there's a lot of evidence.

As for the snake pit that is "searching", we often see attorneys trying to construct searches themselves and the results are always deplorable. As judges have said, this is an area "where angels fear to tread." In order to keep costs down, you need search methodologies constructed by searching experts. And, even then, studies have shown that they will retrieve only 20-22% of the relevant data on the first pass, no matter what methodology they use (keywords or concept searching). You therefore "learn" from the first pass and then do iterative searches. This is the appropriate approach for the producing party, which will comply with both the letter and the spirit of the federal rules.

Clearly, this process would be for a larger case, and less will be done in a smaller case because proportionality (every judge's darling these days) will come into play, as well it should. The smaller the case, the less e-discovery.

The larger the haystack, the harder it is to find the needle. This is the danger of data dumps. And very few recipients are sophisticated enough to find the needle in a data dump. Searching will invariably result in a lot of "false positives," all of which need to be reviewed for relevance and privilege. Attorney review is ALWAYS the most expensive part (by a huge factor) in e-discovery. This is another reason for getting the original volume of data to be searched reduced. 20-22% of 10 GBs will result in much less to be reviewed than 20-22% of a terabyte. And that's the other part of the equation. In the old days (sadly, only five years ago) we were rarely dealing with anything more than gigabytes. Now we deal in terabytes on a regular basis and are anticipating petabytes of data in the near future. The universe of ESI expands daily.

Part of the solution is to have counsel cooperate. This may be wishful thinking, no matter how many judges preach cooperation. More often than not, one side or both are on the warpath and have arrows drawn on a regular basis. Most of the time, if they let their experts talk to one another, the experts will agree on how to proceed (assuming competent experts on both side who want to do a good job for their clients AND hold the costs down). A regular problem is that EDD companies and lawyers both make more money if the volume of responsive data remains large. Processing (by volume) charges and attorney review fees are much higher. So when we see sloppy work or advice, is it due to incompetence or greed? Our anecdotal sense from being involved in so many cases (nothing to back this up with other than our now finely-honed radar) is that it is about 50-50.

All good experts will tell you that they have tried in many cases to steer the client down the right, and cost-efficient path, only to have their advice ignored. It can be very trying - and you worry that the judge in the case will never know that you tried to get the client to do the "right" and cost-efficient thing only to be blown off for reasons that the expert generally can only guess at. When this happens to us, our staff has clear instructions to document the advice given, so that nothing will come back to bite us.

Data dumps are just another way to "hide the ball" which judges uniformly hate. Counsel would be well-advised to avoid this practice, but as the old saying goes, "The easiest way to get rid of temptation is to succumb to it." We predict that sanctions for data dumps are going to spike in the very near future – hopefully, that will impress upon attorneys that courts intend to curb data dumps and punish those who do not honorably discharge their e-discovery duties.

*The authors are the President and Vice President of Sensei Enterprises, Inc., a legal technology, information security and computer forensics firm based in Fairfax, VA. 703-359-0700 (phone)*  
[www.senseient.com](http://www.senseient.com)