

McAleer, Lynn

From: craigball@gmail.com on behalf of Craig Ball [craig@ball.net]
Sent: Friday, December 14, 2007 1:24 PM
To: Dennis Canty
Cc: Kerns, Kevin; Adupre@mccarter.com; Larry J. Gornick; Jaffe, Jonathan; Jim Freebery; rpass@carltonfields.com; cfield@daylightforensic.com; Bill.Adams@fticonsulting.com; McConnell, Stephen
Subject: Re: SEROQUEL: Redaction of Non-paged Multidimensional Data

Ladies and Gentlemen:

During this morning's discussion of production load databases (i.e., the file or files carrying control numbers and metadata for native production between the parties), an example offered was a redacted 150 page Excel spreadsheet. This struck me as an ideal example for readdressing my directive to AstraZeneca that it may not employ redaction mechanisms that serve to corrupt the content or destroy the usability of the data surviving redaction. In the example offered, the 150 page spreadsheet would be TIFFed, redacted and then the unredacted portions would be subjected to optical character recognition, irretrievably injecting corruption of the numerical data and stripping away the formulae and other embedded content essential for usability.

I commented that the example offered exemplified the sort of spoliative action that should not occur--absent express agreement of the parties or explicit order of the court--in a native production. Mr. Adams properly declined to address this issue, noting only that alternative approaches would be burdensome and entail different methods than those contemplated. In short, he didn't say anything to me to encourage discussion or leave the misimpression that it will be an easy issue to resolve.

So, with counsel on board, I'm returning to the hypothetical 150 page Excel spreadsheet.

First, assuming, as is likely, each page doesn't neatly format to 8 1/2 x 11 inches, if TIFFed, the content will flop willy-nilly and uselessly across multiple tiffs. This is the first reason why native production is essential for this data, and I remind you that AZ insisted on native production as a cure for the cancerous problems heretofore.

Second, beneath many cells are formulae that represent the data entered by the user (in that sense, the "core" evidence in the item). These formulae are principally what distinguish a spreadsheet from a word-processed table. Without them, the cell values are merely runes and their rhyme and reason extinguished. TIFFing the data will produce something that entirely lacks these underlying formulae, destroying altogether the *raison d'etre* of a native production.

Third, the data is no longer intelligible as data. Now, it's essentially ink on a page and no more electronically searchable than the first Gutenberg Bible impressed with movable type.

But it gets worse. To this point, the data has been "merely" folded across unnatural dimensions, stripped of its usability and much user-entered information, then rendered so as to eliminate electronic searchability. Only now does AZ redact portions of the information that is entitled to remove. But AZ's obligation is to produce the remaining information (battered and broken as it is) in a reasonably usable and electronically searchable manner. So, AZ proposes to actually or virtually lay the redacted carcass on a scanner and employ optical character recognition to synthesize some semblance of the electronic

12/20/2007

searchability the Rules require. When the data is text, this more-or-less works because spell checkers can catch many absurd constructions. There will be corruption--at least several words on a page will be changed to say something different than the original--but if those words don't happen to be ones key to word search or other needs, some electronic searchability is re-established...more-or-less.

When the data is *numeric*, there are no means to spell check inevitably myopic OCR. Wrong numbers replace right ones and the data becomes wholly untrustworthy. Now the spreadsheet is actually lying to us.

Usability: gone. Searchability: gone. Integrity: gone. Content: affirmatively misrepresented. Time to produce it to the other side.

The contemplated redaction is a miracle drug with one minor side effect: it kills every patient who takes it.

As I expressed in the last call, I do not view it as my role to shepherd methodologies that I know to be faulty and which will ultimately deprive either side or the court of information to which they are entitled under the Rules and the law. The Court is the ultimate arbiter of these matters, but it is my obligation to help the court appreciate the technical issues and their impact on the integrity of the process and the evidence. I'm not content to report that we used industry standard methods to produce useless, corrupted evidence. Redaction in the manner described above merely defers intractable problems to a point where they cannot be readily resolved and thus stands in opposition to the mandate of Rule 1.

I call on the parties to adopt methods that preserve the usability, searchability and integrity of the electronic evidence.

--

Craig Ball
Attorney and Technologist
Certified Computer Forensic Examiner
1101 Ridgecrest
Austin, Texas 78746
TEL: 512-514-0182
MBL: 713-320-6066
FAX: 512-532-6511
E-MAIL: craig@ball.net
WEB: www.craigball.com