





# Litigation in the era of ubiquitous data

## Categorizing data and putting it to use to prove your case

By Nathaniel Leeds

A mentor of mine once told me he loved paper. Why? Approximately: "paper does not lie, it's too dumb to lie, it never expected to be there, it remembers everything, it never covers up, and it always answers the question."

He never met data. He'd have thrown paper over and eloped.

### Who we are; where are we going?

As litigators, we live in interesting times – and behind them. Most of what we do remains as it was in the era of paper: interrogatories, requests for production, Bates stamps, and witness memories. But, society is moving on. Increasingly people live in data, from our phones and cars to our work computers, to the badges we use to swipe into our offices, we leave behind a river of data and that same data often mediates our most intimate human interactions.

I anticipate that over the next 15 years some of us will develop the sophistication to move beyond paper, (and witness testimony for that matter), and start to harvest the data around us to show juries with near-seeming certainty

Copyright © 2017 by the author.
For reprint permission, contact the publisher: **www.plaintiffmagazine.com**




what happened that led to our client's misfortune.

In order to come to terms with the ubiquity of data we need to develop an understanding of how data works, and how to think about its benefits and shortcomings. The goal of this article is to propose a way of categorizing the types of data we all come across to help question it and put it to use.

## Liar's data

One of the first data-heavy trials that I did involved a five-fatality vehicular manslaughter. Our theory was that the defendant-trucker had fallen asleep at the wheel. His claim was that he had suffered an unanticipated diabetic attack.

The public defender offered up the defendant's log books which showed only modest hours behind the wheel in the two days leading up to the accident. My CHP MAIT investigator urged me to subpoena his employer to produce the dispatch logs. As it happened, the trucker was paid by the load, so the company kept meticulous records of when he had driven. The results were obvious: he kept two sets of books so he could work double and triple shifts. My officer explained that this was a common practice.

It is not only a common practice among truckers. "Liar's data" is a byproduct of our data culture. The good part about data culture is that we are able to track all sorts of human activities we want to improve, and keep tabs on. The bad part is that we are constantly asked to explain ourselves.

The more we are asked to record data and report it, the more people tend to lie. It is remarkable how quickly this lying becomes normalized as "Liar's data." The few among this magazine's readership who ever worked for billable hours will quickly understand this phenomenon.

The hallmark of "Liar's data" is that it is data which serves no purpose to the person who is producing it, apart from justifying themselves to someone else. If data serves no purpose to the person making it (like the logbooks in my case), it is susceptible to errors and lies.

I regularly see this in the computer entries nurses make: Many nurses keep the key information they observe during their shift in a notebook some refer to as their "nurse's brain;" the "nurse's brain" is the data they actually use to track their patients. The data they record in the computer system is just a chore.

So, they get sloppy and inaccurate in what they input at the end of the shift. Once you become comfortable with creating sloppy data, you assume others are and that the data is unreliable and – therefore – not useful or important. Once it is not important, it does not feel wrong to create, or manipulate entries.

***TIP:*** Find out what the data you are reviewing means to the person who created it. If it's entirely a chore, and a regular chore which challenges the creator's autonomy and independence, question its accuracy and how often people are "fudging" their entries.

## Narrow data

Whereas "Liar's data" is unreliable because its creators tend to pollute it, narrow data is accurate, but is problematic because it is not complete.

In a roadway defect case, one of the central datasets was from a pavement survey that the defendant city had done. The survey was meticulously carried out by people in the public works department who really cared about the city's streets and wanted to collect the information to target their repairs efficiently. In furtherance of this goal, the city bought an expensive pavement-survey program that ran the NP-complete optimization calculations to figure out how to route equipment to fix the most, worst streets every year.

The problem was that to do the calculations, the program needed really harmonized survey data. So, the program asked that all data be inputted into a computer form. The form did not have any place to observe pavement defects the authors of the program had not figured out how to contend with – like the subsiding shoulder that sent my client tumbling into a $250K medical bill.

Narrow data can be worse than no data when it guides decision-making and can be the foundation for very good policy and procedure liability theories.

***TIP:*** If any data is ever entered in a form which does not allow for, or encourage, free-form text, examine whether there was something about your scenario that a reasonable person would include in a free-form field, if they had only been given a chance.

## Money data

When looking at the quality of data, one consideration is how close the data is to the core interest of the person producing it. Whereas "Liar's data" gets polluted because the producer does not care, and narrow data is incomplete because of a lack of imagination, the data the producer cares about tends to be thorough, and accurate.

For most businesses the most important data is the data they put on their invoices. They will invest considerable time and resources into making sure all of that data is accurately captured and easy to read.

In my practice I spend considerable time with medical billing data, because it is often far more accurate than anything else in a medical record. In hospitals, medications are dispensed by Pyxis machines, which record every pill into the billings system; generic supplies like suture kits have bar-codes which are scanned; bar-code printers mark blood and urine samples and on and on.

In contrast, the official medical chart relies on the accuracy and diligence of doctors and staff, and the Electronic Medical Records can be spread among a choir of uncoordinated legacy software systems: one for the hospital, one for the ER, one for the lab, and so on. The billing data is always simple and accurate.

One of the rhetorical benefits of money data is that jurors tend to resent

Copyright © 2017 by the author.
For reprint permission, contact the publisher: **www.plaintiffmagazine.com**




organizations which are more diligent in their billing than in doing their jobs. This is particularly true in cases where the bills created are bills jurors might have paid.

In a sidewalk-defect case, the city initially claimed that although it had a system in place to do regular inspections, those inspections were not recorded if there was no defect. So, I asked for the parking-ticket and parking-meter data for the block in question. As is often the case, the city might not have kept good records of their inspections, but looking at the money data I could establish that a city employee had walked down the subject sidewalk every other day for the last 20 years – and apparently never thought to tell anyone the sidewalk was crumbling.

***TIP:*** Follow the money.

## Passive data

The holy grail of data collection is the data which is collected and harmonized without us knowing it. I think of this as data 2.0. The first generation of computer systems sought to organize, analyze, and publicize data people had always had as a part of their lives: price data, inventory data, data about when you want a taxi, restaurant reviews. Data 2.0 seeks to collect data we never cared about or shared: our heartrates, how many steps we have taken, when our lights turn on and off, what magazine covers we have scanned, or whether our friends have been to a particular coffee shop.

There's an expression that "if you are not paying for it, you're the product." There is a whole industry designed to collect and sell our data, either to us, or someone else. The implicit exchange is that we will willingly give over the new data and get some benefit. The goal of this industry is to make the arrangement easy: the easier it is for us to give our data, the less we will expect in exchange.

The more passively the data is collected, the less prone to manipulation and the more granular the data is. The hallmark of passive collection is the disappearance of the computer screen and interface – if your Fitbit had a screen, you would never upload your Fitbit data, depriving Fitbit of a potential sales opportunity.

Although our most private data will, with any luck, mostly be shielded from discovery in most cases, the technology and methods that collect personal data will find business applications that create discoverable data that can find its way to the courtroom.

We are working on a roadway case that involves an unwitnessed, fatal solo-bicycle crash. We pulled the data off of the decedent's fitness computer and matched it up with data other riders had posted on the social networking site, Strava, and located witnesses who had seen the decedent shortly before the crash, learned how fast the decedent was going, how hard he was braking, and more importantly, learned how fast the typical rider went on that section of roadway.

This level of data analysis would have been impossible without the passive data collection that is part of many fitness computers. In five years it will be the standard in our profession.

***TIP:*** Look for small data collection devices. The smaller the device, the more likely it is collecting data which can only be viewed when uploaded elsewhere. If it is uploaded for analysis, it is collected and useable.

## The new tool kit: Dealing with bulk

Data is not what many lawyers were trained to do. We are good at thinking on our feet, listening to people, and articulating our positions. Few of us were attracted to litigation because of our overriding love of data analytics.

That is too bad. The world is changing. Attorneys who cannot handle data will be increasingly unable to serve their clients as well as those who can. I once read the transcripts of an attorney who took three PMK depositions without effectively establishing that a troubling lab result was available to a doctor before a patient was discharged from the hospital. This was an expensive and time-consuming comedy of errors: the defense counsel was apparently incapable of locating someone who could talk clearly about the core data issues in the case, and the plaintiff's counsel did not understand the systems well enough to make use of the deponents that were provided.

## There are two tools essential to the data-savvy litigator

First, Excel. Most datasets can be produced in comma-delimited formats which can be quickly dumped into an Excel spreadsheet so they can be harmonized, analyzed and cross-linked with other data in the case. Without some tool to manipulate the data you get, you will never see the patterns and omissions – there is just too much data.

Second, the PMK notice. One of the biggest challenges in any data case is getting the data-novice in opposing counsel's office to find the right person to get you the data you are entitled to. Even in the data-rich world of medical malpractice most defense attorneys are extremely bad at getting and analyzing their own client's data. This is something of a problem of narrow data: there is no billing code for "monkeying around with a spreadsheet."

It has become my practice to send out a PMK notice on the data types I anticipate finding at the same time I send my first round of written discovery. I will typically set the data-PMK deposition 60 days out at the defendant's offices or hospital (so I can look at their computer screen) and explain to opposing counsel that the deposition is being set in the event that they are unable to get their client to produce the data we need for the case. Although the documents provided with the first production rarely include the necessary data, forcing defense counsel to talk to their client and find an appropriate data-PMK often yields results in complete supplemental productions.

Finally, it is worth considering when you want data in your case. As much as

Copyright © 2017 by the author.
For reprint permission, contact the publisher: **www.plaintiffmagazine.com**




I like data, I like liars more. A good case is one in which you have data to prove what happened. A great case is where you have a deposition transcript of the defendant lying, and data to prove it. Because most opposing counsel is not data-savvy, it can sometimes be tactically advantageous to hold back on data discovery until the defendant's key witnesses have had a chance to lie, and then push for data to unravel the lie.


Leeds

*Nathaniel Leeds handles a broad range of civil cases on behalf of consumers and small businesses for* Brent, Fiol & Pratt LLP, *including personal injury, medical malpractice, and business litigation. He started his legal career as a Deputy DA in Merced County where he tried numerous cases to a jury, including third-strike felonies, juvenile sexual assaults, and manslaughter. He brings this extensive trial experience to his civil practice. He is a University of Chicago graduate. He received his law degree from UC Hastings in San Francisco.*

Copyright © 2017 by the author.
For reprint permission, contact the publisher: **www.plaintiffmagazine.com**