



# The Hidden Bazinga of Dirty Data

Topic: *Legal Data Intelligence*  
Assoc. Keyphrase: *Data Management*

Summary:

*What can you do when the hidden Bazinga of dirty data plays tricks on your company's bottom line? More than you think ....*

[Listen with Spotify](#) | [Reprint Permissions](#)

§ § §

It's a solitary evening in unit 4A of the Los Robles Apartment Building in Pasadena, CA.

Or so it seems.

Our hero, Leonard, sits dejectedly on the couch, contemplating his fractured love life. His posture, the look on his face—they tell us everything we need to know. This day could not get any worse.

Or so it seems.

Suddenly, in a spasm of prank-filled fury, his roommate Sheldon bursts out from under the couch cushions in zombie attire, and screaming with reckless abandon. Leonard leaps from the couch clutching his heart. The zombie roommate is

pleased, and he sums up the moment with these unforgettable words:

“Bazinga, punk.”

If you're a sentient being in North America, you already know that's a scene from the popular sitcom, [The Big Bang Theory](#), and that “Bazinga” was the trash-talking catchphrase Dr. Sheldon Cooper used to punctuate his reckless mischief. What you may not know is that the biggest Bazinga-maker in 21st Century business is this:

[Dirty Data](#).

## The Hidden Bazinga of Dirty Data

If dirty data lived in the Los Robles Apartment Building in Pasadena, CA, the Bazingas would be coming hard, fast, and often. Consider this sampling of the havoc it already creates for business:

- » IBM estimates that dirty data is a cause of more than \$3 trillion in US business losses every year.
- » One-third of America's business leaders “don't trust the information they use to make decisions.”
- » Harvard Business Review reports that knowledge workers waste 50% of their time “hunting for data, finding and correcting errors, and searching for confirmatory

sources for data they don't trust."

» According to a report from Experian, "On average, U.S. organizations believe 32 percent of their data is inaccurate, a 28 percent increase over last year's figure of 25 percent."

» MIT Sloan Management Review estimates that dirty data will cost your company between 15%-25% of your gross revenue this year.

Unlike *The Big Bang Theory*, these kinds of data disasters don't deliver comedic effect. Dirty data is—literally—costing employees their jobs, putting companies at risk, and significantly hindering growth in the US economy every year.

## A Law Firm Example

The real problem with dirty data is that it's not just one problem. It's a culmination of problems that occur when [data management](#) is not optimized. As an example, let's take a look at a fictional company we'll call "Big Money Law Firm," or BMLF.

BMLF is launched in 2008 as an under-the-radar legal firm specializing in real estate law and intellectual property. After a rocky start that coincided with a global recession, they've been riding a wave of success and growth ever since. Before long they'd expanded to healthcare law, corporate law, financial transactions, and more. Like the legal services industry as a whole, BMLF has seen their revenue increase every single year since 2009. Always profitable, their industry has exploded in the last decade, generating \$256 billion annually in the United States alone—and over \$889 billion worldwide. In 2010, BMLF earned \$69.8 million in revenue. By 2020, their annual revenue had grown to over \$390 million, and this formerly "small" firm boasted six distinct divisions, expansive offices in three major metropolitan areas, and over 400 total employees.

It's a rosy situation, right? Well, you'd think so, except:

» In line with the [Experian report](#) mentioned above, 32% of BMLF data is dirty, which is a 28% increase over last year's figure of 25%. The reasons for that include: duplicate data records, inaccurate data, non-integrated data, business rule violations in data, and inconsistent data management.

» Similar to the experience of Jun Wu, a data scientist working in investment banking, BMLF has completed five different mergers and acquisitions in the last 10 years—but trying to integrate legacy data structures from those M&A operations left them with more than 1,000 disparate systems that they have yet to successfully consolidate. The result? Data degradation, information inconsistency, multiples sources of diverging "truth," and major decisions based on inaccurate data.

» BMLF has siloed data management within its six divisions, meaning bad data in one division gets passed on to the next in line, corrupting data function in the second division. That process continues until no one knows for sure which data can be trusted, and which should be thrown out.

» Congruent to the [MIT Sloan](#) reference above, this year alone BMLF's dirty data has cost them 22% of potential revenue—over \$85 million. Poor data management strategies have wasted corporate resources, decreased corporate productivity, damaged the company's brand in public circles, eliminated opportunities for new business, dumped lost money into useless marketing and communications spends, and prompted layoffs in one unprofitable division.

"Wait a minute," you say, "Big Money Law Firm still earned \$390 million last year. That's pretty good, right?"

Well, that depends. Would you rather your [law firm](#) earn \$390 million in revenue, or \$475 million? Would you rather your company endure layoffs, or continue hiring? Would you rather your company brand be known for excellence, or mediocrity? Would you rather ... well, you get the idea.

And that's why we call dirty data the "hidden Bazinga" of business. It steals from you that which is significant—but which is frequently unseen. It costs a company more than most are willing to acknowledge.

## Expert Strategies for a Bazinga-Free Data Environment

The good news is that your company doesn't have to suffer through Bazinga madness for long. In fact, you can take meaningful steps to [reduce the impact](#) that dirty data has on your business this year, and beyond. Here are three expert strategies to get you started.

### 1. Connect the Beginning with the End

Thomas C. Redman advises this as your first step: "Connect data creators with data customers."

The popular author of *Getting in Front of Data Quality*, he says, "From a quality perspective, only two moments matter in a piece of data's lifetime: the moment it is created and the moment it is used ..."

"Improving data quality isn't about heroically fixing someone else's bad data. It is about getting the creators of data to partner with the users—their 'customers'—so that they can identify the [root causes](#) of errors and come up with ways to improve quality going forward."

### 2. Play Small Ball

[Kyle Williams](#) is director of business consulting at data intelligence firm, Blue Margin Inc.—and he’s a St. Louis Cardinals fan. He recommends taking a page out of the MLB playbook when it’s time to face off against dirty data:

“Play small ball.”

In baseball, the “small ball” offensive strategy focuses on situational hitting and on-base exploitation to make runs add up through the innings. So, instead of pursuing a power-hitting, home-run-or-bust approach, the team executes simple, inside-the-park tactics that lead to victory, such as the bunt, sacrifice fly, hit-and-run, and steals. A veteran of corporate data intelligence, Williams has seen the same strategy work well in mid-market corporations.

“Doing nothing is essentially just waiting for the world to change and the reality to change within a business,” Williams says. “Playing ‘small ball’ in your data strategy allows you to pick up what’s accessible and reportable, and start to develop a strategy that supports key business areas.

“Small ball allows momentum to be gained for an organization, focusing on certain operational or financial areas of the business, and getting traction within the organization to develop BI. It’s the catalyst for implementing new systems or processes to leverage either the current team, or the current data system, to make it better.”

### 3. Implement Scorecards—and Let Everyone See Them

Nikki Chang was tasked with improving data performance for the massive operations of the drilling department at Chevron, a \$230 billion energy behemoth. Her solution? Data scorecards tied to bottom-line metrics—available anytime to everyone.

Chang’s first-year goal was a minimum of 95% accuracy in all new data for drilling wells. She upped the standard to 100% for the second year. Then she measured the performance of those goals in real time, on a daily basis—tracked on a scorecard that everyone could see.

The results were astounding, and almost immediate. Some teams started daily reviews to quality check their data. Other rig groups created an ongoing competition, challenging each other to exceed company goals to win. Within only eight months, 13 of Chevron’s 15 affected divisions (86%) had already achieved the Year 1 goal, and the two remaining divisions were close as well.

Chang points to the easily-accessible scorecards as the catalyst for this considerable change. “Everyone can see how they’re doing at all times,” she says. “This is important—when they try to improve something, they get to see whether or not they were effective. And they can see how they’re doing relative to their peers.”

Dirty data is hiding in your company, but you don’t have

to let it scare you. You can take control of your data—and let it drive you toward success instead.

## Three Key Thoughts:

1. “Dirty data is—literally—costing employees their jobs, putting companies out of business, and significantly hindering growth in the US economy every year.”

2. “The real problem with dirty data is that it’s not just one problem. It’s a culmination of problems that occur when data management is not optimized.”

3. “Improving data quality isn’t about heroically fixing someone else’s bad data. It is about getting the creators of data to partner with the users—their ‘customers.’”

## For Further Reading:

- » [Legal Services HQ™](#)
- » [Data Wins Championships](#)

[Schedule a demo with a Blue Margin data expert](#)



### Sources:

“The Apartment Building.” The Big Bang Theory Wiki. [https://bigbangtheory.fandom.com/wiki/The\\_Apartment\\_Building](https://bigbangtheory.fandom.com/wiki/The_Apartment_Building)

The Big Bang Theory, season 5 episode 7, “The Good Guy Fluctuation,” directed by Mark Cendrowski, aired October 27, 2011, on CBS. [https://www.imdb.com/title/tt2082016/?ref=fn\\_at\\_tt\\_1](https://www.imdb.com/title/tt2082016/?ref=fn_at_tt_1)

IBM, “The Four Vs of Big Data.” [https://www.bluemargin.com/hubfs/Current%20Assets%20\(2021+\)/05a%20Blog-Images-TTD/blog-2021-4-Vs-of-big-data%20-%20IBM-1.jpg](https://www.bluemargin.com/hubfs/Current%20Assets%20(2021+)/05a%20Blog-Images-TTD/blog-2021-4-Vs-of-big-data%20-%20IBM-1.jpg)

Thomas C. Redman, “Bad Data Costs the U.S. \$3 Trillion Per Year,” Harvard Business Review online. <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>

“The Cost Of Dirty Data - Regit,” YouTube video, 2:25, “Regit,” May 9, 2017, <https://www.youtube.com/watch?v=rOzyKMmUUG4>

Thomas C. Redman, “Seizing Opportunity in Data Quality,” MIT Sloan Management Review online, <https://sloanreview.mit.edu/article/seizing-opportunity-in-data-quality/>

E. Mazareanu, “Legal services industry in the U.S. - Statistics &

Facts.” Statista, May 9, 2019, [https://www.statista.com/topics/2137/legal-services-industry-in-the-us/#dossierSummary\\_chapter1](https://www.statista.com/topics/2137/legal-services-industry-in-the-us/#dossierSummary_chapter1)

E. Mazareanu, “Size of the legal services market in the United States in 2019 and 2020, by category” Statista, Oct 23, 2019, <https://www.statista.com/statistics/741393/size-of-the-legal-services-market-by-category-us/>

Erin Haselkorn, “New Experian Data Quality research shows inaccurate data preventing desired customer insight,” Experian online, <https://www.experian.com/blogs/news/2015/01/29/data-quality-research-study/>

Jun Wu, “What is Dirty Data?” Towards Data Science online, <https://towardsdatascience.com/what-is-dirty-data-d96abbd254e>

“Dirty Data,” Technopedia online, <https://www.techopedia.com/definition/1194/dirty-data>