

# Why the “Let’s Migrate Everything!” Mentality Is a Bad Idea

By Jeff Fehrman, Chief Strategy Officer of Mindseye

When your organization upgrades a system, or switches vendors, your IT team follows a specific process that probably includes the following steps: installation, configuration, testing, migration, and training. The migration piece is important — that’s where you ensure that your existing data is ported over into the new system. All of it.

But is that really such a good idea? It actually may not be. The truth is, any system is likely to have some bad data that should either be fixed or purged — but definitely not migrated.

## What makes data bad?

In *The Bad Data Handbook*, author Q. Ethan McCallum defines bad data as data that “... eats up your time, causes you to stay late at the office, drives you to tear out your hair in frustration. It’s data that you can’t access, data that you had and then lost, data that’s not the same today as it was yesterday ... In short, bad data is data that gets in the way.”<sup>1</sup>

Some data is bad for technical reasons; values are missing, data is incorrectly formatted, or records are duplicative. This is common in databases and applications.

Some data is **dark data**, or data created during regular business activities that the organization fails to use for other purposes. This is more likely to happen with unstructured files,

such as documents, spreadsheets, presentations, etc., on file servers.

And some data is simply junk with no business value; this is prevalent in email systems.

## Bad data is bad for business

Bad data is a big problem for enterprises. Ovum Research reported that poor data quality costs businesses at least 30% of revenues. According to Gartner, 31% of data-related initiatives failed last year due to bad data, and another 24% experienced project delays due to data-related problems.

And yet, too many organizations take a “let’s keep everything” attitude when it comes to their data. One factor that contributes to this mentality is the plummeting cost of data storage. From an IT perspective, it seems easier and cheaper to keep it than to figure out what should be disposed. The same goes when it’s time to migrate data to a new or upgraded system.

Storage isn’t the only cost to consider, though. [The 1-10-100 rule as applied to enterprise data](#) posits that verifying the quality of a record costs \$1, cleansing and de-duplicating a record costs \$10, and working with a record that’s never been cleansed costs \$100. That means that bad data costs the organization 100 times more than good data.

## A more discriminating approach requires technology

Cleansing data isn’t easy — if it was, there wouldn’t be so much bad data out there. [A recent InformationWeek article](#) noted that “migration is an opportune time to think about data quality and to ensure the new system is fed with the highest quality data it can be supplied with.”

As the article notes, there are data quality tools that can automate data profiling in advance of migration to uncover problems that will affect the migration process and reduce the effectiveness of the new system.

This is only part of the story. After the migration, the data won’t stay “clean” — all the factors that contribute to making good data go bad will still be in play. To address this, enterprises can also leverage technology on an ongoing basis to classify data.

Maintaining data classifications can improve data quality by categorizing data for better management and “findability” and eliminating unnecessary data. Data that meets the defined criteria gets automatically classified, eliminating the need to do it manually.

It does take time in the beginning to train a system and develop the classifiers. You don’t want to “boil the ocean,” but a migration project is, in fact, a perfect starting point.

## Stop perpetuating the problem

By continuing to save and migrate bad data, enterprises compound the problem — and the impact — of bad data. But, by leveraging technology in a smart way and using migration as an opportunity to systematically attack the problem, organizations can improve the quality of the data that they manage and store.

## End notes

<sup>1</sup> McCallum, Q. Ethan, *The Big Data Handbook*, O’Reilly (2012)

### MINDSEYE SUPPORTS EFFECTIVE AND REUSABLE CLASSIFICATION

*The Mindseye platform is a powerful and comprehensive eDiscovery solution that can help you classify and categorize information and enable you to:*

- Quickly cull junk data and uncover relevant information through a sophisticated, iterative keyword search
- Inform future cases by reusing/repurposing workflows and historical data
- Integrate your eDiscovery process with your information governance strategy
- Use built-in analytics to refine and improve your information governance strategy and develop data classifications

*For more information about the Mindseye platform or to request a demo, please*

[www.mindseyesolutions.com](http://www.mindseyesolutions.com).



Since joining Mindseye, Jeff has helped the team design and launch Mindseye’s appliance and mobile offerings and established key technology partnerships. Jeff is on the Board of Governors for the Organization of Legal Professionals (OLP) and is an active member in the Sedona Conference and AllIM. He frequently speaks on innovations and obstacles facing corporations and law firms today. He has developed classes approved by the Continuing Legal Education (CLE) boards in the states of California, New York, Pennsylvania, and Virginia and has trained hundreds of attorneys in the technical aspects of digital forensics and electronic discovery. Jeff is co-founder of the *EDD Blog Online*, one of the electronic discovery industry’s premier resources for news and related content, with over 4,000 articles to date, and co-author of *Focused Discovery: Perspectives on the Future of eDiscovery and Information Governance*. Contact Jeff at [jfehrman@mindseyesolutions.com](mailto:jfehrman@mindseyesolutions.com) or (571) 483-0639.

**MINDSEYE**

**Discover More. Review Less.®**