

DATA

# Bad Data Costs the U.S. \$3 Trillion Per Year

by Thomas C. Redman

SEPTEMBER 22, 2016



Consider this figure: \$136 billion per year. That's the research firm IDC's estimate of the size of the

**Lead your way to success with HBR**  
**Subscribe now >**

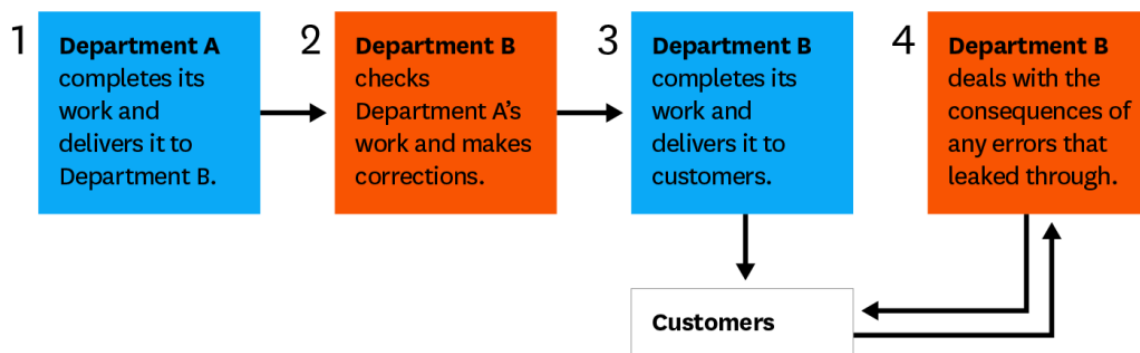
While the numbers are not really comparable, and there is considerable variation around each, one can only conclude that right now, improving data quality represents the far larger data opportunity. Leaders are well-advised to develop a deeper appreciation for the opportunities improving data quality present and take fuller advantage than they do today.

The reason bad data costs so much is that decision makers, managers, knowledge workers, data scientists, and others must accommodate it in their everyday work. And doing so is both time-consuming and expensive. The data they need has plenty of errors, and in the face of a critical deadline, many individuals simply make corrections themselves to complete the task at hand. They don't think to reach out to the data creator, explain their requirements, and help eliminate root causes.

Quite quickly, this business of checking the data and making corrections becomes just another fact of work life. Take a look at the figure below. Department B, in addition to doing its own work, must add steps to accommodate errors created by Department A. It corrects most errors, though some leak through to customers. Thus Department B must also deal with the consequences of those errors that leak through, which may include such issues as angry customers (and bosses!), packages sent to the wrong address, and requests for lower invoices.

### The Hidden Data Factory

Visualizing the extra steps required to correct costly and time-consuming data errors.



SOURCE THOMAS C. REDMAN

© HBR.ORG

I call the added steps the “hidden data factory.” Companies, government agencies, and other organizations are rife with hidden

**Lead your way to success with HBR**  
**Subscribe now >**

Such hidden data factories are expensive. They form the basis for IBM's \$3.1 trillion per year figure. But quite naturally, managers should be more interested in the costs to their own organizations than to the economy as a whole. So consider:

- 50% – the [amount of time that knowledge workers waste](#) in hidden data factories, hunting for data, finding and correcting errors, and searching for confirmatory sources for data they don't trust.
- 60% – the estimated fraction of time that data scientists spend cleaning and organizing data, [according to CrowdFlower](#).
- 75% – an estimate of the fraction of total cost associated with hidden data factories in simple operations, based on two simple tools, [the so-called Friday Afternoon Measurement and the “rule-of ten.”](#)

There is no mystery in reducing the costs of bad data – you have to shine a harsh light on those hidden data factories and reduce them as much as possible. The aforementioned Friday Afternoon Measurement and the rule of ten help shine that harsh light. So too does the realization that hidden data factories represent non-value-added work.

To see this, look once more at the process above. If Department A does its work well, then Department B would not need to handle the added steps of finding, correcting, and dealing with the consequences of errors, obviating the need for the hidden factory. No reasonably well-informed external customer would pay more for these steps. Thus, the hidden data factory creates no value. By taking steps to remove these inefficiencies, you can spend more time on the more valuable work they *will* pay for.

Note that very near term, you probably have to continue to do this work. It is simply irresponsible to use bad data or pass it onto a customer. At the same time, all good managers know that, they must minimize such work.

---

**Lead your way to success with HBR**  
**Subscribe now >**

work diligently to find and eliminate the root causes of error. Those that follow this regimen almost always reduce the costs associated with hidden data factories by two thirds and often by 90% or more.

I don't want to make this sound simpler than it really is. It requires a new way of thinking. Sorting out your requirements as a customer can take some effort, it is not always clear where the data originate, and there is the occasional root cause that is tough to resolve. Still, the vast majority of data quality issues yield.

Importantly, the benefits of improving data quality go far beyond reduced costs. It is hard to imagine any sort of future in data when so much is so bad. Thus, improving data quality is a gift that keeps giving – it enables you to take out costs permanently and to more easily pursue other data strategies. For all but a few, there is no better opportunity in data.



Thomas C. Redman, “the Data Doc,” is President of Data Quality Solutions. He helps companies and people, including start-ups, multinationals, executives, and leaders at all levels, chart their courses to data-driven futures. He places special emphasis on quality, analytics, and organizational capabilities.

---

**This article is about DATA**

**+** FOLLOW THIS TOPIC

Related Topics: OPERATIONS

---

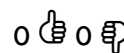
**Lead your way to success with HBR**  
**Subscribe now >**

**Marco Rubio** a year ago

Dear Mr. REDMAN, I would be interested in purchasing your paper but I don't understand if it is the same article written in the above section or a much more comprehensive and detailed study with numbers and figures. I would be interested in understanding how you achieved similar results looking at your conclusions. Please answer me so I can decide how to procede. kind regards

MARCO

REPLY



▼ JOIN THE CONVERSATION

---

#### POSTING GUIDELINES

We hope the conversations that take place on HBR.org will be energetic, constructive, and thought-provoking. To comment, readers must sign in or register. And to ensure the quality of the discussion, our moderating team will review all comments and may edit them for clarity, length, and relevance. Comments that are overly promotional, mean-spirited, or off-topic may be deleted per the moderators' judgment. All postings become the property of Harvard Business Publishing.

---

**Lead your way to success with HBR**  
**Subscribe now >**