



Hidden Costs of Duplicate Customer Data

Excerpted from Dan Power's blog posted on 12/13/2009

Recently a client asked me what percentage of duplicate data was “normal” in customer master data.

My initial response was that, when companies don't have formal master data management, data governance or data quality initiatives in place, duplication rates of 10 to 30 percent (or more) are not uncommon.

In fact, when I was at Dun & Bradstreet, we would routinely see that level of duplication in clients' customer files.

A good example of this was cited in a healthcare field study about how Children's Medical Center Dallas¹ engaged an outside firm to help clean up their duplicate data:

“Solving both the current and future problems around duplicate records helped Children's improve the quality of patient care and increase physician acceptance of the new EHR. The duplicate record rate was initially reduced from 22.0% to 0.2% and five years later it remains an exceptionally low 0.14%. The 5 FTEs initially tasked with resolving duplicate records have been reduced to less than 1 FTE.

For the Children's Medical Center, the results were heartening, not only from a care delivery standpoint but also because of the significant cost-savings that can be realized. A study conducted on Children's data showed that on average, a duplicate medical record costs the organization more than \$96.”

So it is possible to significantly reduce the duplication rate through careful analysis and the application of the right tools, as part of an ongoing data governance program. Even the hospital cited above (and hospitals are usually not mentioned as practitioners of best practices) was able to maintain a duplication rate of only 0.14 percent after five years.

And there are very real costs associated with not cleaning / de-duplicating your customer data. Depending on the functional area (marketing, sales, finance, customer service, etc.) and the business activities you undertake, high levels of duplicate customer data may cause negative effects on your business, including:

- The same communications sent to the same customer multiple times undermines their confidence in your company and impacts your brand,
- Additional marketing expenses (for duplicate mailings and other communications),
- Increased costs for the manual reconciliation of data,
- Resistance by employees (and sometimes customers too) to new systems,
- Multiple sales people, sales teams or collectors contacting the same customer

The best studies I've seen of the cost of duplicate data have been in the healthcare industry. One study² I recently read stated:

“According to Just Associates, the direct cost of leaving duplicates in an Master Patient Index database is anywhere from \$20 per duplicate to several hundred dollars. The lower cost reflects the organization's labor and supply costs to identify and fix the record while the higher expense reflects the costs of repeated diagnostic tests done on a patient whose previous medical records could not be located.

¹ Source: <http://www.justassociates.com/documents/studiesinsuccess.pdf>

² Source: http://www.fortherecordmag.com/archives/ft_081505p30.shtml



The American Health Information Management Association (AHIMA) estimates that it costs between \$10 and \$20 per pair of duplicates to reconcile the records. If the records aren't reconciled, however, the costs are even higher."

Several other case studies support the 10 to 30 percent range I quoted:

- Once the analysis was complete, Sentara discovered they had a significant duplication rate, over 18%. They had attempted to address the duplication rate in the past through a remediation process, but due to either technology issues or because the cost of merging and cleaning up the duplicates across their many different systems was too high, they had not yet successfully reduced their duplication rate. Source: http://www.initiate.com/customers/success_stories/Pages/SentaraHealthcaresNewEMPIReducesDuplicationCreationRates.aspx
- Emerson Process Management faced a tremendous challenge four years ago in getting its CRM data in order: There were potentially 400 different master records for each customer, based on different locations or different functions associated with the client. "You have to begin to think about a customer as an organization you do business with that has a set of addresses tied to it," says Nancy Rybeck, the data warehouse architect at Emerson who took charge of the cleanup. Working with Group 1, Rybeck analyzed the customer records for similarities and connections using everything from postal standards to D&B data, and managed to eliminate the 75 percent site-duplication rate the company suffered in its data. "That's going to ripple through everything," she says. Source: <http://www.destinationcrm.com/Articles/Columns-Departments/REAL-ROI/How-to...ensure-CRM-data-quality-44095.aspx>
- Problem: Number of duplicate records: 20.9% of Utah Statewide Immunization Information System records. Impact of Problem: Difficult to find patients in system—key barrier to provider participation, risk of over-immunization—unable to find reliable patient record, cost of unnecessary immunizations, risk of adverse effects on patients. Source: http://health.utah.gov/phi/brownbag/handouts/2008/USIIS_april.pdf

And here's a good quote from a white paper titled "*Data Quality and the Bottom Line*" by The Data Warehousing Institute:

"Peter Harvey, CEO of Intellidyn, a marketing analytics firm, says that when his firm audits recently 'cleaned' customer files from clients, it finds that 5 percent of the file contains duplicate records. The duplication rate for untouched customer files can be 20 percent or more."

Every organization needs its own metrics but if left unchecked, the duplication problem is a hidden cost that drags at your company, slowing down your processes and making your analyses less reliable.

Sales analysis reports must be accurate and provide a "single customer view" for each of your largest customers. If they don't, the sales figures for those customers are most likely incorrect and then the entire report becomes suspect.

I'd like to end with a great quote on data quality by Ken Orr from the Cutter Consortium in "*The Good, The Bad, and The Data Quality*":

"Ultimately, poor data quality is like dirt on the windshield. You may be able to drive for a long time with slowly degrading vision, but at some point, you either have to stop and clear the windshield or risk everything."

Hub Solution Designs, Inc.

188 Whiting Street, Suite 3, Hingham, MA 02043-3844 USA

office 781.749.8910 | fax 781.735.0318 | www.hubdesigns.com | info@hubdesigns.com

Read More by Dan Power: blog.hubdesigns.com