Workbook

CALCULATING COSTS

The Price of Dirty Data

BY PAUL A. STRASSMANN

TOOL: Why Messy Records Hurt the Wallet

Where do you begin if you want to take advantage of the potential benefits from transforming your systems through consolidation, conversion to a service-oriented architecture, and imposition of enterprisewide rules that would ensure interoperability and security of applications, externally as well as internally? As a first step, you need to adopt standard definitions for metadata—the data about data—to guide the processing of all data inputs, regardless of whether they come from legacy or already transformed applications. The worksheet below is designed to show the expenses associated with having incomplete, inconsistent and inaccurate records in multiple databases—that cannot be easily integrated—at a bank. This example also assumes dramatically reducing the number of databases and applications in use—because that's the key way to reduce errors.

INSTRUCTIONS: Start with the number of data sources in your organization; review samples to estimate completeness and accuracy of data. Follow directions described at left, and fill in your own numbers under "Your Company." To get an interactive version of this worksheet, see: GO.BASELINEMAG.COM/JUL06.

		EXAMPLE	YOUR COMPANY
	BASICS		
	Number of applications	185	
	Number of databases	40	
	SCOPE OF DATA MANAGEMENT		
A	Data sources, e.g., terminals, magnetic card recorders or counters attached to sensors	22,800	
B	Median number of data elements (names, ID numbers, serial numbers of devices, etc.) entered per transaction	8	
C	Median number of daily transactions per source	800	
D	Data elements entered into system per year **(A x B x C x 365)**	53,260,800,000	
	MEASURES OF DATA QUALITY		
E	Completeness of data entry, based on samples taken from suspended transactions or error registers. Example: a phone number that's missing an area code.	97%	
F	Accuracy of data, based on samples compared to templates from a corporate data dictionary. Example: A misspelled name or address.	98%	
G	Duplication of data. Example: John Smith makes two reservations, so his name shows up twice in a reservation database.	5%	
H	Conformity of database entries to defined business rules for security, authentication, etc.	90%	
	METRICS FOR DATA MANAGEMENT		
I	Data requiring corrections per year **(D x ((1 - E) + (1 - F))**	2,663,040,000	
J	Incorrect additions to database per year **((D x G) + D x (1 - H))**	7,989,120,000	
	COST OF MANAGING DATA INTEGRITY		
K	Data audit and remediation. Based on cost for automated intervention per error.*	$0.10	
L	Administrative cost per defect	$0.08	
M	**Annual cost to manage defective data ((I x K) + (J x L))**	**$905,433,600**	

*The cost of fixing defects goes up as they become rarer, as quality increases. For example, fixing a defect at 99.9999% quality is harder than fixing a defect at 99.5%