

The Cost of Lost Data

The importance of investing in that "ounce of prevention"
[David M. Smith, Ph.D.](#)

Application: The cost of lost data from computers is substantial. Businesses must be proactive in protecting this important resource.

The Nature of the Problem

All computer users are familiar with the problem of lost data. Fortunately, most such incidents are relatively inconsequential, representing only a few minutes of lost work or the deletion of unnecessary files. However, sometimes the nature of the lost data is critical, and the cost of lost data is substantial. As reliance on information and data as economic drivers for businesses continues to increase, owners and managers are subject to new risks. One study reports that a company that experiences a computer outage lasting for more than 10 days will never fully recover financially and that 50 percent of companies suffering such a predicament will be out of business within 5 years.^[1]

Levels of Risk

Of course, the value of lost data varies depending on their application, as well as the potential value that can be captured from use of the data. The loss of computer code, for example, represents a significant loss of value because computer code must be rewritten by highly skilled and highly paid software developers. In contrast, the loss of a customer history database would represent a less significant episode of data loss, assuming original source copies of the information are available. In this case, although the data would need to be re-keyed, it could be done by lower skilled and lower paid data entry personnel.

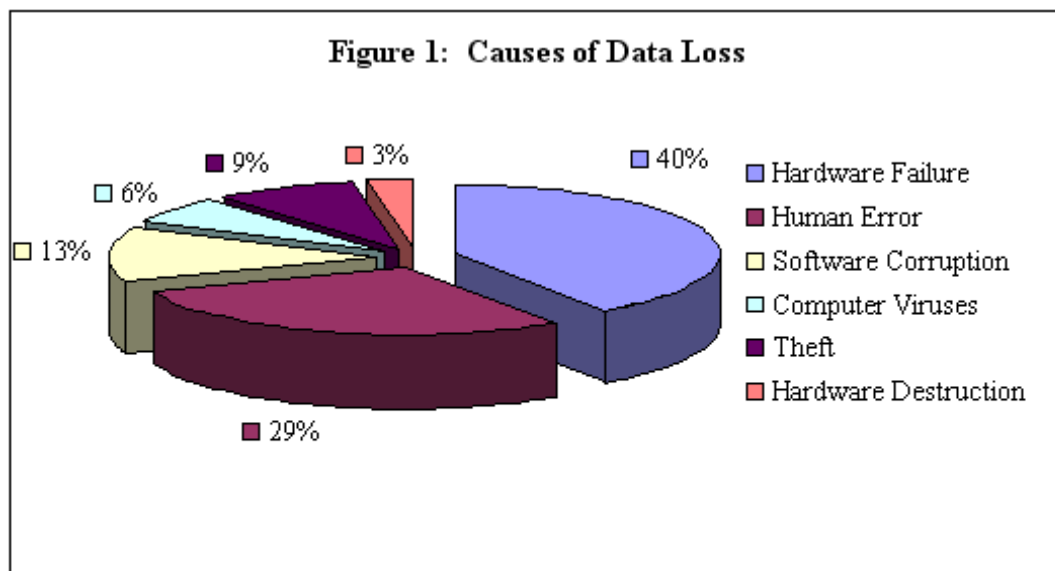
Using available data sources, this paper attempts to quantify the costs associated with episodes of data loss in the aggregate for the US economy. Implications of these findings for business owners and managers will also be discussed.

PCs in Use

Companies increasingly rely on data in a distributed environment. Therefore, the examination of data loss here will focus on the level of the personal computer.^[2] US businesses use an estimated 76.2 million PCs to aid in the production of goods and services. Laptops are relied upon more and more, with a current installed base of 15.2 million units, or about 20 percent of all business PCs. The number of desktops in use currently totals approximately 61.0 million units.^[3]

Episodes of Data Loss

Statistics on data loss are sparse. Data loss incidents can be hardware- or software-related. Consequently, a consideration of both is necessary to estimate the magnitude of data loss. Thus, this study combines two data sources to estimate the magnitude of data loss in the US: (1) claims data from an insurance company that insures computer hardware; (2) survey data from a company that specializes in data recovery.^[4] Estimates from this combination suggest that the most common cause of data loss is hardware failure, accounting for 40 percent of data loss incidents. These include losses due to hard drive failure and power surges. Human error accounts for 29 percent of data loss episodes, which include the accidental deletion of data as well as accidental damage done to the hardware, such as damage caused by dropping a laptop. Software corruption, which might include damage caused by a software diagnostic program, accounts for 13 percent of data loss incidents. Computer viruses--including boot sector and file infecting viruses--account for 9 percent of data loss episodes. Theft of hardware, especially prevalent with laptops, accounts for 6 percent of data loss incidents. Finally, hardware destruction, which includes damage caused by floods, lightning and fire, accounts for 3 percent of all data loss episodes. The relative magnitudes of the different types of data loss are illustrated in Figure 1.



Source: Author's estimates based on data from Safeware, The Insurance Agency, Inc., "2000 Safeware Loss Study," 2001; and ONTRACK Data International, Inc., "Understanding Data Loss," 2003.

These data may be mapped to census ("installed base") data on computers to estimate the number of severe data loss episodes that occur each year. Table 1 reports the results of this mapping, estimating 4.6 million episodes of severe data loss per year. Reflected in these data are significant differences in the incidence of data loss between laptops and desktops. While less than two percent of desktops are likely to experience an episode of data loss each year, the corresponding rate for laptops is greater than ten percent.

Causes of Data Loss	Episodes of Data Loss
Hardware Failure	1,849,800
Human Error	1,345,300
Software Corruption	588,600
Computer Viruses	294,300
Theft	403,000
Hardware Destruction	126,100
Total	4,607,100

Source: Author's estimates based on data from Computer Industry Almanac, 2003; U.S. Dept. of Commerce, National Telecommunications and Information Administration, *A Nation Online: How Americans Are Expanding Their Use of the Internet*, February 2002; John Spooner, *News.com*, "Laptops gain in PC Market," August 20, 2001; Safeware, The Insurance Agency, Inc., "2000 Safeware Loss Study," 2001; and, ONTRACK Data International, Inc., "Understanding Data Loss," 2003.

The Cost of a Data Loss Incident

An episode of severe data loss will result in one of two outcomes: either the data are recoverable with the assistance of a technical support person, or the data are permanently lost and must be rekeyed.^[5] A calculation of the average cost of each data loss incident must take into account both possibilities. The ability to recover data depends on the cause of the data loss episode. The permanent loss or theft of a laptop whose data have no tape backup will result in permanently lost data. In addition, fire or flood damage can also make the possibility of data recovery very remote. For other causes of data loss, data recovery specialists are becoming more adept at restoring inaccessible data.^[6] Taking into account all causes of data loss, evidence suggests that in 83 percent of the cases, data may be recovered.^[7]

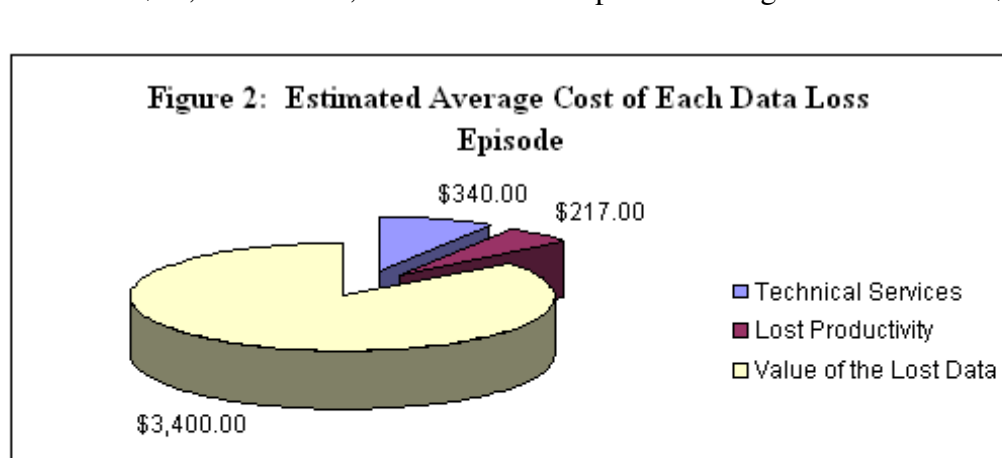
The first cost of data recovery to be considered is that associated with hiring a computer support specialist in the recovery effort. If there is a computer support specialist employed within the company, both the number of hours needed to recover the data and the cost of employing this individual must be taken into account. Most recent information from the Bureau of Labor Statistics states that the average computer support specialist earns an estimated \$28.10 an hour, including both salary and benefits.^[8] The time needed to recover data may vary greatly. If a data backup exists and is readily accessible, the time needed to recover data may be very short. At the other end of the spectrum, if the data are corrupted on the hard drive, several days may be required to retrieve the data.

If the average time needed to recover lost data is approximately six hours, the cost of using an employee to recover lost data is approximately \$170. However, if a firm does not employ a specialist who is able to retrieve lost data, the company must go to an outside firm to attempt data recovery. Outside data recovery specialists can be much more expensive than in-house sources, sometimes exceeding two to three times the cost of an in-house specialist. Thus, taking into account that an outside specialist must often be used in data recovery attempts, one can conservatively estimate the minimum cost of outside technical support to recover lost data to be around \$340.

During the time in which the attempt to recover data is underway, an individual is unable to access his or her PC, thereby reducing productivity, which in turn impacts company sales and profitability. This opportunity cost--lost productivity due to computer downtime--impacts a company's income statement just as do other more common and explicit costs. Lost productivity represents missed opportunities for income generation. Some employees are directly involved in sales and revenue production; others are involved in more supportive or indirect roles. Economics teaches that each employee's productivity, or contribution to firm revenue, can be approximated using the individual's compensation.^[9] Available data sources suggest that individuals who use computers at work earn an average of \$36.20 an hour in wages and benefits.^[10] Thus, \$38.70 for six hours totals approximately \$217.^[11]

The final cost to be accounted for in a data loss episode is the value of the lost data if the data cannot be retrieved. As noted earlier, this outcome occurs in approximately 17 percent of data loss incidents. The value of the lost data varies widely depending on the incident and, most critically, on the amount of data lost. In some cases the data may be re-keyed in a short period of time, a result that would translate to a relatively low cost of the lost data. In other cases, the value of the lost data may take hundreds of man-hours over several weeks to recover or reconstruct. Such prolonged effort could cost a company thousands, even potentially millions, of dollars.^[12] Although it is difficult to precisely measure the intrinsic value of data, and the value of different types of data varies, several sources in the computer literature suggest that the value of 100 megabytes of data is valued at approximately \$1 million, translating to \$10,000 for each MB of lost data.^[13] Using this figure, and assuming the average data loss incident results in 2 megabytes of lost data, one can calculate that such a loss would cost \$20,000. Factoring in the 17 percent probability that the incident would result in permanent data loss, one can further predict that each such data loss would result in a \$3,400 expected cost.

Added together, the costs due to technical services, lost productivity, and the value of lost data bring the expected cost for each data loss incident to \$3,957. (See Figure 2.) It should be noted that most data loss incidents (approximately 83 percent) result in much lower average costs (\$557), but in the smaller portion of cases in which the data are permanently lost, the average costs are estimated to be much higher (\$20,557). In addition to highlighting the significant costs involved in re-keying data, these figures reflect the importance that data played in creating value for businesses. Once data are lost, those value-creating opportunities are also lost. These losses are multiplied in a networked environment. A survey conducted in 2001 by Contingency Planning Research reports that the majority of companies estimate the average cost of computer network downtime to exceed \$50,000 an hour, and for some companies that figure rises to over \$1,000,000 per hour.^[14]



Source: Author's estimates based on data from Denise Deveau, "Lost all your data? Time to Call the Experts," *The Globe and Mail*, February 25, 2000; Bureau of Labor Statistics, *Employer Costs for Employee Compensation*, March 2003; and Bureau of Labor Statistics *Occupational Employment Statistics Survey*, 2001.

Total Annual US Data Loss Costs

When information on data loss episodes is mapped along with the cost data, an estimate of aggregate data loss may be obtained. This calculation, reported in Table 2, estimates that annual data losses to PCs cost US businesses \$18.2 billion.^[15] This estimate represents an increase from a 1999 study that estimated the annual cost of lost data to be \$11.8 billion.^[16] Although it is difficult to measure with precision the cost of lost data, and the analysis is sensitive to the assumptions that underlie its calculations, there are several reasons to believe that \$18.2 billion is a conservative estimate. First, that figure does not take into account costs that are difficult to quantify, such as lost sales and reputation damage a firm may experience during an extended period of computer downtime. In addition, research in the field of network economics suggests that extra costs would be incurred if a data loss incident occurs to two or more PCs on a network. Such additional cost is due to the interdependence and reliance that each computer user experiences when working with other computer users. As noted earlier, research on incidents of server downtime suggests that such costs can be significant. Finally, it is important to note that these figures do not include any collateral costs that may be incurred in some instances of data loss, such as when damaged hardware must be replaced.

Types of Loss	Average Cost of Each Data Loss Incident
Technical Services	340
Lost Productivity	217
Value of the lost data	3,400
Sub Total	\$3,957
Total US Data Loss Costs	\$18.2 Billion

Source: Author's estimates based on data referenced in all prior tables and figures.

Trends and Implications

What trends are likely to impact the potential for data loss in the future? Available evidence suggests that PC users are more likely now than ever before to use power surge protectors and virus protection software.^[17] In addition, Safeware, a company that specializes in insuring PCs, reports that computer thefts appear to be declining as a percentage of computer loss incidents.^[18] This is positive news. However, it is the opinion here that two trends will drive the annual amount of data loss upward: (1) increased reliance on laptops, which are much more likely to suffer episodes of data loss than are PCs, particularly from accidental damage due to dropping; and (2) more data stored in smaller spaces, since hard drive capacity continues to increase. Conservative estimates place the rate of data growth at 80 percent per year.^[19] Not only is the amount of data increasing, but business reliance on data is also rising.^[20]

Implications from this research are clear. Business managers should invest in technologies that can reduce the possibility of data loss.

Examples include the use of computer virus detection and back-up systems. PCs should be password protected, to reduce the value of a stolen PC to a potential thief. Also serving as a theft deterrent are computer-tracking services which serve as a sort of "LoJack for laptops."

However, even in the face of strong protection measures, some episodes of data loss will inevitably occur. Plans to deal with such episodes can mitigate recovery times. And although back-up protocols are common for server-located data, plans to protect data in a distributed environment are less commonplace.^[21] Since the technologies available to back-up data are often reasonably priced, cost does not necessarily present a stumbling block in preventing permanent data loss. A simple and essentially zero cost data back-up procedure involves copying data on CD writable disks using the pre-loaded software that comes with PCs. IT staff should hold training sessions on such protocols because adherence to such procedures will rely on individual users following through with such protective procedures. A saying that precedes the advent of the computer is appropriate here: an ounce of prevention is worth a pound of cure.

Endnotes

1. Jon Toigo, *Disaster Recovery Planning: Managing Risk and Catastrophe in Information Systems*, (Yourdon Press, 1989).
 2. Data, as defined here, are the bytes that reside on personal computers. Data could be more broadly defined as all digital media, but such a consideration is beyond the scope of this study. See Simon Forge, JP Morgenthal, and Richard Pak, *Manager's Guide to Distributed Environments: From Legacy to Living Systems*, (New York: John Wiley & Sons, 1998).
 3. Data in this paragraph from the *Computer Industry Almanac*, 2003; U.S. Dept. of Commerce, National Telecommunications and Information Administration, *A Nation Online: How Americans Are Expanding Their Use of the Internet*, <http://www.ntia.doc.gov/niahome/dtr/>, February 2002; and John Spooner, *News.com*, "Laptops gain in PC Market," August 20, 2001.
 4. See Safeware, *The Insurance Agency, Inc.*, "2000 Safeware Loss Study," 2001; and ONTRACK Data International, Inc., "Understanding Data Loss," 2001, (2003).
 5. This assumes that the vast majority of computer users are unable to recover from a severe data loss incident without the assistance of a technical support individual. The household equivalent of this assumption would be that the vast majority of households could not recover from a severe plumbing problem without a plumber. This also assumes that the original source information is available in hard copy or some other form from which it can be rekeyed.
 6. The Data Recovery Group reports that they are able to recover inaccessible data in 95% of incidents. See <http://www.datarecoverygroup.com>, (2003). It is noted that data recovery firms may have an incentive to overestimate their success rates.
 7. An 83 percent recovery rate is reported by Denise Deveau, "Lost all your data? Time to Call the Experts," *The Globe and Mail*, February 25, 2000.
 8. From Bureau of Labor Statistics, *Employer Costs for Employee Compensation*, March 2003, and *Occupational Employment Statistics Survey*, (2001).
 9. The reasoning is fairly intuitive: a company will pay an employee as long the individual adds to firm profits. The company will stop paying an employee when the revenue generated from that individual is exactly equal to the compensation paid.
 10. The average compensation for white collar workers as reported in *Employer Costs for Employee Compensation*, Bureau of Labor Statistics, (March 2003).
 11. It may be claimed that an individual could move on to other tasks, which in turn would only reduce their productivity by a fraction of this amount. However, it is common that the PC user must work closely with the data recovery specialist in the recovery effort. In addition, productivity could be hampered for several days if the computer must be sent to an outside specialist.
 12. A National Computer Security Association (now TrustArc Corporation) survey reported that for an average engineering department it would cost \$100,000 to rebuild 20 megabytes of data.
 13. For example, see Stuart Hanley, "Keep Those Data Protection and Recovery Options Open," *Storage Management Solutions*, November 1997; and ONTRACK Data International, Inc., "The Data Recovery Solution," (1998).
 14. See Contingency Planning Research, 2001 Cost of Downtime Survey, <http://www.contingencyplanningresearch.com/2001%20Survey.pdf>, (2002).
 15. This could be considered the annual data loss estimate for the year 2003. However, although this study aims to utilize the most up-to-date data sources available, some data are from years prior to 2003. Thus, \$18.2 billion represents the best estimate of annual data loss, based on the most recent sources available.
 16. David Smith, "The Cost of Lost Data," *Storage Management Solutions* No. 4 (1999): 60-2.
 17. TrustArc Corporation, *7th Annual ICSA Labs Virus Prevalence Survey*, (2002).
 18. Safeware, *The Insurance Agency, Inc.*, "2002 Safeware Loss Study," (2003).
 19. Jon William Toigo, "Storage Disaster: Will You Recover?," *Network Computing*, (March 5, 2001).
 20. Toigo, 2001.
 21. Forge, Morgenthal and Pak, 1998.