# Protecting Your Library's Digital Sources

## The Essential Guide to Planning and Preservation

**MIRIAM B. KAHN**

**American Library Association**
Chicago 2004

While extensive effort has gone into ensuring the reliability of information appearing in this book, the publisher makes no warranty, express or implied, on the accuracy or reliability of the information, and does not assume and hereby disclaims any liability to any person for any loss or damage caused by errors or omissions in this publication.

Composition by ALA Editions in Palatino and Optima using QuarkXPress on a PC platform

Printed on 50-pound white offset, a pH-neutral stock, and bound in 10-point cover stock by McNaughton & Gunn

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992. ∞

**Library of Congress Cataloging-in-Publication Data**

Kahn, Miriam (Miriam B.)
Protecting your library's digital sources : the essential guide to planning and preservation / Miriam B. Kahn.
p. cm.
Includes bibliographical references and index.
ISBN 0-8389-0873-X
1. Libraries—Safety measures. 2. Archives—Safety measures. 3. Library information networks—Security measures. 4. Electronic records—Conservation and restoration. 5. Emergency management—Planning. 6. Computer security. 7. Data protection. 8. Electronic data processing—Backup processing alternatives.
I. Title
Z679.7.K385 2004
025.8'2—dc22 2003023834

Copyright © 2004 by the American Library Association. All rights reserved except those which may be granted by Sections 107 and 108 of the Copyright Revision Act of 1976.

Printed in the United States of America

08 07 06 05 04 5 4 3 2 1

*To My Family*

# Contents

# Preface

Today libraries, archives, and organizations of every size are creating websites with digital materials that they maintain day in and day out. Some repositories are digitizing written works in the public domain and putting them on the Web and on CD-ROM (or equivalent technology) to make them accessible to the public, with the twofold aim of (1) providing access to the information today while preserving the fragile or rare originals for the future, and at the same time (2) wanting to "keep" the digital information for the future. Other institutions have created massive digital databases of their holdings, including books, manuscripts, and images that they want to be accessible on the Web. On top of all this is the incredible amount of office files, materials, circulation records, and cataloging records that are created every day.

The very largest libraries and archives in the world, in conjunction with national library and archive consortia, have been looking at the issue of the preservation of digital materials for several years. They have begun publishing "best practices" that any library of any size can incorporate into its project planning, budgets, and follow-up. Corporations and financial institutions have focused their energies on contingency plans and enterprise-wide backup of data.

Less well-funded libraries and archives and other cultural institutions need a practical "how-to" guide to plan for the future of their data, whether it be for access tomorrow, next year, or in ten years.

This book will look at the prevention of loss, the restoration of data or digital materials, and planning for long-term access to these materials. There are two major issues in planning to prevent the loss of digital materials: physical loss of data and hardware; and intellectual loss. Section 1 of this book will look at that common disaster, loss of data from a system or server crash, and how to prevent the loss of data through regularly scheduled backups. It will examine procedures to plan for the restoration of hardware and software after a loss and the salvaging of physical storage media and hardware. Section 2 of this book will discuss procedural decisions to increase the longevity of digital materials, documents,

images, and databases. It will look at the digital preservation options of copying, migrating, reformatting, and converting. (Just because a library or archive stores information on a diskette, CD/DVD, or magnetic tape does not ensure that the data will be readable in the future, even as soon as tomorrow.) The discussion of project planning will include questions about what types of storage formats to consider, types of operating systems, software, hardware, "refreshing of data," and that all-important planning and budgeting for future supplies, storage, staff, and expertise to maintain the viability of digital materials into the future.

Chapter 10 of this book provides 29 different checklists to help digital project managers and information technology specialists formulate a disaster response plan and consider a variety of issues in planning for the long life of their projects.

# Acknowledgments

Thanks to Cynthia Read-Miller and Jane Kenamore for ideas about individuals and their need to back up data. Thanks to Clark Searle for his help with insurance and the disaster response industry. Special thanks to Nancy Birk, former special collections librarian and university archivist at Kent State University, and to Julie Callahan, librarian at Columbus Metropolitan Library in Columbus, Ohio, for listening to my ideas about disaster response for libraries. Thanks to Tom Benjamin at Iron Mountain. Thanks to Robin Fremer who helped me through the rough days and Wayne Luff, my business partner, who helped me make time in my schedule to write. And most of all, thanks to my excellent editor, Marlene Chamberlain, for her editorial support and encouragement throughout the writing process.

# Introduction

Despite the events of September 11, 2001, the power blackout in the northeastern United States and Canada in August 2003, hurricanes, tornadoes, and other disasters natural and otherwise, one out of every three businesses or organizations has no disaster response plan for digital resources and will not survive a failure of its computers or technology.[1] What we need are plans for and forethought on how to keep our operations functioning despite loss of data and computers. Creating plans to deal with a loss of computer data is essential to the future of libraries, archives, museums, and their parent institutions. Without that data, service cannot be provided, be it local or remote, and your patrons will seek information and assistance elsewhere.

In the same way, planning for the retention of digital materials is an essential part of protecting and ensuring the long life span of digital materials. It is also the key to an effective disaster response or contingency plan for computer operations. While libraries, archives, and museums have been writing or thinking about disaster response plans for years, they often do not think about the computer operations in their own buildings, let alone in the institution as a whole. What libraries, archives, and museums have usually considered is what to do if their audiovisual materials get wet in the course of a disaster. By audiovisual materials, I mean audio and cassettes both digital and analog; CDs in all their manifestations, DVDs, videodiscs, etc.; and LPs and their earlier forms. Some institutions may also consider photographs, negatives, microfilm, and microfiche to be audiovisual materials. To this end, the institutions consider how to determine whether the audiovisual materials are wet, and if they are, whether it is cost-effective to dry them out. In the case of all audiovisual formats, there are two questions to be answered: (1) is the object dry and clean so that a user can access its information; and (2) is playback equipment available either at the cultural institution or at the patron's home so that they can play the medium? These same issues apply to digital materials.

We have learned from the audiovisual field the importance of proper storage conditions,

both physical and environmental; the importance of playing and copying the contents of the storage medium to ensure that "new" playback equipment can "read" whatever is on the tape or disc; and that some organization or institution should take responsibility for keeping the "last" piece of playback equipment so that future librarians and archivists can access the medium. The audiovisual preservation field has been very successful in its efforts to retain physical storage media and our access to them.

So now we must apply these lessons to digital materials. Digital materials come in many forms today and will continue to increase in number and complexity with each succeeding year. The professional literature describes digital materials in a number of ways, most often as either "born digital," that is, the data was never in any format but digital; and digitized or scanned, that is, the data existed in another format, print or three-dimensional, and is now accessible in some digitized format. In a true sense, born-digital materials are the most vulnerable to time and technological improvements because they were never anything else. Digitized materials are now in their new format for a variety of reasons, most often to provide access to the image or the intellectual content without further damaging or putting more stress on the original object. It is generally understood that the original object is not available to the casual user, just those who need to work with it for scholarly or research purposes.

So how does all this relate to protecting data? In a nutshell, if we don't protect the physical object from deterioration due to poor storage conditions and environment, then it will be impossible to access the data stored on that medium in the future. And if we don't take into consideration how we will access the data when the hardware, software, and operating systems no longer exist, then all that hard work is lost to future scholars. Data in the form of letters, reports, statistics, images, databases, catalogs, and full-text books, as well as every website that has ever been mounted and modified, are all at issue and vulnerable to time and technological change.

As previously stated, this book is divided into two major sections. The first looks at protecting data and the physical object or medium from "normal," everyday losses from fire, water, careless backups, and computer or server crashes. In essence we will create a disaster prevention and response plan for computers and their data, as well as look at an overarching contingency plan for recovering the operations of the computer services and related departments in your institution. The second section will look at current practices for preserving digital materials and will define the methods used for "refreshing" data. We will also look at the issues of how future librarians, archivists, and computer technicians will access the data on current storage media, how the profession will try to ensure that it will be there later, and how we will pay for it.

The issues we will consider in this publication are different from those in most books entitled "preservation of digital materials" or "preserving digital materials." For the most part, these books discuss how to organize a project to digitize materials from print or three-dimensional formats to some digital format. They are not necessarily looking at digitization as a means of preserving the object from excessive handling and abuse or as a long-term storage and access issue.

Advertisers, the general public, and the promoters of these processes often use the word "archival" to imply preservation without differentiating these two terms. When talking about technology, it is important to understand what these terms mean colloquially and what they mean in the library/archive and information technology (IT) fields. When you *archive* data, it is stored in a retrievable, accessible manner for the long term. When you *preserve* data, the physical object or storage medium is stored for the long term under stable environmental conditions, but the data may not be readable or accessible. First we will explore the issue of keeping the data and its physical storage medium so that

your organization can get to it if computers crash or your building is inaccessible. Then we will look at creating digital materials and storing them in an archival manner so that each item is retrievable in the future.

## NOTE

1. Business First "Survey: Firms Unprepared for Operations, IT Outages," *Business First,* March 7, 2003, A27.


Why is planning for the loss of data so important? For a number of reasons, including the amount of time it takes to reconstruct the data the way it was originally created, accessed, or used. Think about how much time it takes to load your software and data onto a new computer. It takes many hours. What if you lost the data and software as well as the computer? Now think about how long it would take to order a new computer with all the features you want, get it delivered, purchase the software, install it, and then try to find your latest data backups. If you are like most people, you don't have all your data backed up, or you've only backed up older versions, original renditions, last year's data, or those special projects you could never live without. Worse yet, you might have backups for the data, but the only versions of the software you have are older and so you cannot read the files because software is never forward-compatible. And if you thought this could never happen to you, imagine what it would be like if everyone in your organization lost their data all at the same time, and each department had its own sense of how important its job is to the mission of the institution. But if you do some advance planning and are proactive about it, then you and the computer department will have a good idea of what is needed and where and when each function is slated for recovery.

There are several names for plans that you may have read about in "disaster response" or business literature: contingency planning, business continuity planning, emergency management plans, and disaster response, to name a few. Each aims to do the same thing: get operations back up and running.

"Disaster response plan" is the most common term used by libraries, archives, and museums. These plans deal primarily with recovering damaged books, audiovisual materials, photographs and paintings, and three-dimensional objects. While making decisions about the recovery of objects, the disaster response plan also discusses how to get both public services and behind-the-scenes operations back up and running.

Contingency plans deal primarily with computers and all related data services, such as fiscal operations and data transfer.

Business continuity plans look at how to keep the business as a whole running: operations, personnel, financial, and customer services.[1] The word "continuity" is used because if there are a multitude of plans at an institution, it is essential that the disaster response team members be aware of the other plans and their priorities for restoring services, and that the disaster response team leaders communicate with each other as they design, coordinate, and carry out their plans.

Emergency management plans usually fall within the jurisdiction of an emergency management agency and are discussed in conjunction with damage or destruction to a city or county infrastructure. Such efforts are often organized by law enforcement agencies in a state, city, or county and focus on restoring the operations of the government, public services, and infrastructure, especially after a natural disaster such as a tornado, flood, or hurricane.

Each of these types of plans has its trade literature and sources for specialized information. Industry and planning information is found in the American trade journals *Disaster Recovery Journal* and *Contingency Planning and Management* and in the British publication *Survive.* Of course, you will find articles about disaster prevention and its many permutations in business journals, library/archive and records management journals, and almost every other type of publication. The World Wide Web is also a great place to look for general information, the names of organizations and service providers in your geographic area, and basic guidelines for plans.

As opposed to just "thinking about" disaster response planning for recovery of your three-dimensional objects and then designing a plan on the fly, you must be proactive about planning for the recovery and long-term retention of your data. *If* you do not plan ahead, then you are bound to lose what has been so painstakingly created, stored, and preserved.

## NOTE

1. Jeffrey W. Greenberg, "September 11, 2001: A CEO's Story," *Harvard Business Review* (October 2002).

# 1

# Preventing Common Causes of Loss

*It will never happen here!*

We've all heard this expression, but since September 11, 2001, the attitude it represents seems to be less of an issue. The further we get from that date, however, the more the denial syndrome will come back. Planning for the loss of data, and for retaining information over long periods of time, are the two best methods of ensuring that "it will never happen here"!

## Where Libraries and Archives Are Today

Libraries and archives have been developing disaster response plans for quite a while now. This doesn't mean that every institution has one, but the number is growing, especially in the aftermath of 9/11. Institutions are looking at the need to protect their collections from any type of damage or loss. Unfortunately, libraries and archives are all over the board when it comes to actually having an up-to-date written disaster response plan.

When we look at disaster response plans for computer services and the electronic information maintained by libraries and archives, we again find that the preparedness levels of institutions vary. The degree to which a library or archive is computerized plays a large part in determining how prepared the organization is for the loss of access to computers and online data. Some electronic resources are backed up onto institution-managed servers and networks; some are not. Indeed, some of the most valuable data is stored on local hard drives and is backed up irregularly by the individuals responsible for those computers. This is the first issue of concern in this book: the protection of data and computers today, so that libraries and archives can function when computer systems fail.

There are three items to consider when examining how to protect your data for the short and long term. The first is *backing up the data* onto a removable storage medium. Item two is *storing the data in a safe, secure location.* The third item is *being able to recover the data* and *reinstall it onto the existing or re-created databases* your organization maintains.

It is important to realize that backing up data onto tapes protects the organization from loss of information as a whole. However, if you want to retrieve a *specific* piece of information, you must provide some indexing to go along with the data. In this way, you can retrieve the data without having to reload the entire tape. This is the difference between backing up data and an archival database. When you think about creating backup files, you might want to think about the long-term issues at the same time. For a little more time and money, you can accomplish both tasks at once, and save serious expenses over the long term.

## Common Causes of Loss of Data

### *Accidental Erasure*

Accidental erasure is probably the most common type of loss of data. This happens when you close a file and don't save it, or write over the original file when you meant to save it as

something else. You might mistakenly delete a file, but if you don't do something else in between, and you don't turn off the computer, you can usually restore the file. Now if you erase or delete a file accidentally from a network, then you would need to have the computer systems department restore the file. They can only do this if the file has been backed up when the server is backed up. If you lose a file that you were responsible for backing up on your own hard drive, then you may be out of luck, unless it was on your backup tape. Anything you don't want to lose should be backed up onto some external medium (diskette, CD-ROM, tape, flash memory, etc.).

You can lose your current changes when there is a power surge or your computer crashes and you forgot to hit the save button. In this case, the computer might remember that the file was open and present the "last saved version" of the file. But if you haven't backed up the file since the day before, and you lose the current work, you may be out of luck and have to create all the changes again. It is this loss that is most costly for individuals and organizations to re-create. Worst of all, insurance policies don't cover this type of loss. (This issue is discussed in detail in chapter 6.)

Many computing departments leave the backup of individual data to the individual user. Unless your users are compulsive, they won't back up their own data regularly. One solution is to have all users' directories reside on the network so that data is backed up automatically when the system is. The only limitation here is the amount of memory storage capacity allocated to each person or account. Copies of individuals' data should be kept both in their offices and some other remote location, such as their car, home (if permitted), or the data center itself. Some mechanism needs to be in place to check that the storage medium is accessible, readable, and contains data. Storage media should be labeled with information that includes what type of data the medium contains, when the backup or the data was created, the programs used to create it, and whose data it is.

### *Viruses and Worms*

Computer viruses and worms invade computers through unprotected computers via e-mail, diskettes, and CDs, where the unwitting or unsuspecting receiver opens an attachment. The most common computer virus invades your e-mail program, causing your computer to send out messages with the same attachment to everyone in your address book. Most of the time, you have no idea that your computer is sending out viruses until a friend lets you know. Other computer viruses attach to program files, causing the programs to perform improperly, while others infect DOS files, corrupting the operating system and requiring many hours of work to reformat the hard drive. You need to have your operating software on a bootable disc to regain control of the computer from the virus.

The best protection against unwanted viruses is to install antivirus software on your computer and your network. Once you install the software, it will check your hard drive, e-mail, and all files every time you open them. You can also check disks and CDs before installing or playing them. The antivirus software will update itself regularly and should be upgraded annually. Antivirus software companies include Norton and MacAfee.

Computer software companies and computer news services post fixes for viruses that you can download and install. The virus fix will remove the virus. In some cases, you will lose all your e-mail and addresses, especially if the virus is invasive and the cure is to reformat your hard drive. Back up your address book and bookmarks or favorites regularly using the export feature in your e-mail and browser. If your institution doesn't back up e-mail to tape banks, then back up your e-mail using the archive function on your e-mail program. Be careful to archive to separate disks or files, since the archive feature can write over existing data files.

Other computer viruses are called worms. Worms corrupt your hard drive by copying their files over and over again onto it until there is no space left. Once there is no space left on your hard drive, computer programs won't function.

Removing the worm includes reformatting the hard drive and restoring all the programs and files. If you haven't backed up your files and programs, you will have to re-create them, a most time-consuming chore.

Accidental erasures of data, viruses, and worms aren't the only types of loss that can occur, however. In today's technology-dependent world, the loss of power and of phone, cable, and Internet connections, let alone access to the computer itself, can become a major crisis.

## *Power and Telecommunications Outages*

### POWER OUTAGES

On August 14, 2003, just after 4 p.m., a blackout hit the northeastern United States and Canada, turning off computers, lights, and everything else from Long Island and Manhattan Island to Detroit, Toledo, and Cleveland and north into Canada as far as Toronto and Ottawa. The blackout brought airlines and trains to a halt; gasoline stations couldn't pump gas; some cities had no water, and restaurants and groceries did their best to distribute perishable food before it spoiled. Some businesses had backup power for 4–6 hours, others for 12 hours; some had diesel generators, but the majority of businesses just shut down. The blackout lasted anywhere from twelve hours to four days depending on where you were in the country. Some businesses had contingency plans so that their current data was automatically backed up to remote locations or generators kicked in, and their staff reported to an alternative operations site to begin repairs and the restoration of service.

To prevent the wholesale loss of data, an uninterruptible power supply provides battery backup to allow for time to close computer systems down systematically rather than have them crash and users lose their active data files. Some of today's programs will actually save files that are trapped on the computer due to system failure or power outages, but they only remember the "last saved version." When the system is restored, the computer system will ask if you want to use the last "lost" version. Of course, you could set your computer to save regularly, especially if you are using stand-alone equipment. The main problem with auto-backup is that it will overwrite the current file again and again. If you are editing that file to become something else, as when you modify a template, the computer will overwrite the original version, and then you won't have the original in the backup. Develop a routine, if you are editing an existing file, to immediately save the file under a new name. In this way the auto-backup doesn't overwrite data that you wish to retain.

A large institution should have its computing systems running on at least two different power grids. That way if one grid goes down, there is another available for running backup servers and communications.

### TELECOMMUNICATIONS OUTAGES

Loss of telecommunications is very frustrating in today's working and researching environment. Libraries and archives today access their external databases and information resources while hardwired to a T1 line (or something faster), ISDN, fiber-optic cable, or a cable connection. The internal network may be wireless or hardwired together and can include intranets, circulation systems, online catalogs, and other internal databases that are not accessible from outside computer systems. Very small libraries and archives may still be dependent upon dial-up service providers.

Regardless of what type of telecommunications system your institution employs, you should have some type of backup service for your communications needs. This goes for phone service as well. Some institutions have a few non-electric phones to use should the electricity go out; that way they can dial out for assistance.

Institutions need to think about some alternative access method to web-based and server-based systems when high-speed telecommunications connections are down, disrupted, or disabled. The alternative access could be by dial-up or another cable system. When reviewing your

disaster response plan, you should think about using two different companies or access methods. But beware: all the telecommunications companies in your area may be using the same optical fiber network, so you may need to establish an agreement with a neighboring institution. You should investigate this thoroughly before signing any contracts.

Other alternatives for telecommunications may include contracting with an "on demand" wireless communications network that will provide you with e-mail and paging should the telecommunications in your area go down.[1] Of course, there is a caveat with this option. If there is a major disruption in the infrastructure in your city, as happened in New York City on September 11, 2001, then bandwidth for wireless communications may be jammed or unavailable for emergency contracts. If your institution thinks wireless communications are essential during a disaster or` loss of computer access, then you should think about a contract with a company that provides the "on demand" wireless service.

#### SUSTAINED LOSS

No matter how you plan to continue providing services during a power outage, should your institution suffer a telecommunications outage that results in loss of contact with the computer systems, then you must have an emergency plan in place to forward all phone numbers and lines to an alternative location. You must also have personnel at the alternative location to answer the phones. This is true of the telecommunications methods for accessing servers and networks in your institution. If you need to provide different access methods to the data over alternative data lines, then you need to write this into your plan, and establish a contract with a company that can provide the service. (Your computer disaster response team will need to get another mirror site operational or your original site operational as quickly as possible. Be certain to build in backups to the mirror site.)

If there is a sustained loss of telecommunications or power, your institution should declare a "disaster" with your data storage/backup service provider and activate your disaster response plan for computer services. This would mean gathering the backup software and data and mounting it on computers and servers in a remote location. This facility can be near your facility or in another city.

After September 11, 2001, some businesses experienced a reluctance on the part of their employees to travel any distance from home. This resulted in the increased use of mobile recovery centers for data disaster recovery. When your organization declares a "disaster," all that practicing and testing by the information technology and systems department for the full restoration of data will come in handy. This will also be the time to offer your information and research services from a remote location and take queries by phone, fax, and e-mail instead of in person.

## Regular Backup Procedures—What's So Important?

To forestall the loss of data and loss of patrons, let alone work product, regular backups are imperative. This isn't the only procedure to incorporate into daily routines. You must check that the backup system is functioning, both the hardware and the tape, and that the data and associated software are readable and usable. Otherwise, you will end up having to create the data again.

The whole idea behind data and program backup is to prevent loss of data from the computer system when the server or network crashes. Unlike disaster response planning for the physical items in a library or archive, you cannot create a response and recovery plan on the fly. Careful premeditated steps must be taken to ensure the safeguarding of data and computer systems. To this end, computer disaster response plans focus on the prevention of loss first and foremost.

When an organization loses its data, the financial implications can be enormous. One study examined the cost of lost data and concluded that "every cent of data backup is worth

$2500 of data re-entry."[2] Another cost study reported that employees on average cost $36 per hour; imagine the amount of productivity lost should your department, building, or institution lose its ability to access computer services for even one hour.[3]

Time studies show that for every hour of data lost, it takes one day to reenter it within the course of a normal workday; and for every day of lost data, it will take a week to reenter the data. Just think about the types of work you perform in the course of a normal day, then multiply that by the stress of performing work under less than normal conditions in a strange environment. Now add to this the loss of your data files and the need to re-input data and re-create original ideas, projects, catalogs, and websites. The reconstruction of data is time-consuming and extremely expensive, and many times can be avoided.

### *Types of Backup: What Method to Use*

Business literature discusses two important issues to consider when selecting backup methods: the time it takes to recover the data and how much data loss you can live with; and the time it takes to get the operations back up and running.[4] Or to think about this another way, what is the cost to your institution of using or having old data, and what is the cost of being off-line? Can your institution survive with old transactional data? Do you even have any that is mission critical?

What type of backup method do you choose for your institution, building, network, or office? Well, it depends upon the resources (people, equipment, and money) your organization has, and how critical the data and the computer operations are to the survival of your institution. Two questions to ask the computer disaster response team are the following ones. How long can the institution, library, or archives do without computers? How long can you do without data before the amount of data to input is crippling? (Business calls these two factors RTO, or recovery time objective—how long you can live without the application—and RPO, or recovery point objective—what is the maximum amount of data you can lose?)[5]

In the case of financial institutions, their data is so critical that even a few hours of downtime could destroy their cash flow, a crisis from which they might not recover. In the case of a library, archive, or museum, data is not quite as critical, although the institution as a whole might be more data-dependent. So what type of backup method do you choose? Well, if you need the data *now* no matter what the crisis, then your organization should be using data mirroring or replication. Data mirroring means that for every keystroke made, a mirroring keystroke is made in a remote location. So the data are almost identical. This type of backup is commonly employed by financial institutions, the travel industry, and other data-critical sectors. Other businesses use a fairly continuous transfer or backup of data, so the data might be shunted to another location every few minutes, or every hour. These incremental backups can be retrieved quickly but are more difficult to synchronize. Smaller businesses, archives, and libraries use data vaulting, in which the data is sent via the Internet, FTP (file transfer protocol), or TCP/IP (another type of transmission protocol) in regular cycles, but perhaps as slowly as every night. Tape and disc backup are still commonly used, store the least information, and take the most amount of time to back up, but are the least expensive to maintain. Tape and disc backup also mean that data is less critical to the survival of your institution, since reinstallation of the software and data can take between 12 and 72 hours, whereas data mirroring is almost instantaneous. No matter which method you choose for your institution, you must test the backup tapes to confirm that they are recording data and that the hardware is working.

Business continuity planners remind us that crisis management is better than disaster recovery, especially when dealing with loss and recovery of data and computers. If you have a plan and have backed up your data, the total cost to the institution of recovery is decreased by 25–50 percent.[6]

#### INCREMENTAL BACKUP

The most common backup method for individuals to use is the incremental backup. Only the files that have changed are copied onto some type of removable storage medium or a remote data storage site. These files are usually projects that people are working on or are critical data that the individual doesn't want to lose. At some point in time, the individual will back up everything, usually when they are switching to a new computer. Unfortunately, if the computer contracts a virus or a worm in the meantime, all the data that was not backed up recently, or ever, may be lost. Individuals should consider taking the time to back up their data on a regular basis, say once a month. That way what was a dormant project will not be lost.

Some automated backup programs also back up only what has been changed since the last time. This works well as long as a complete backup is performed on a regular basis, at least once a week. However, the organization runs the risk of losing all the data that was created since the last full backup. If the system that fails is your circulation or cataloging system, your staff may not be able to re-create the lost data.

Complete daily backup is the best method for smaller institutions that have automated systems, providing they can afford the loss of that day's data.

#### DATA MIRRORING

Data mirroring or replication includes synchronizing two or more data servers with exactly the same information. So with every keystroke or transaction you make, the same exact transaction is made to the second or third database. There are several types of replication that can be employed. The first is to back up or replicate on a second drive in the same cabinet. This would ensure against physical loss of or damage to the original drive. The second type of replication would be to have another drive or server in a different machine or power grid. This level of replication is common on a large campus. The third and highest level of replication is across geographical areas. This is also described as "real-time" transaction replication. Some businesses in New York may replicate their data on the West Coast, or in Europe or Asia. The only caveat is the time it takes for the transaction to transmit from one location to another. The farther away the transaction has to go, the longer it takes, so these same organizations might have one local remote location and one far away. This is the type of real-time backup that is used by the financial industry and airlines.

Backup procedures always require scheduling, and checking that a backup was successful and the data is readable is an essential part of testing the backup.

### *Backup Media: What Format to Pick*

Now that you have decided that you must back up your data and the method to be used, you will need to decide what removable or remote storage medium to use. There are several choices depending upon the amount of data you have, the frequency of backup, and how critical the data is to the survival of your institution.

Diskettes, CD-ROMs, and flash memory cards are great for the local backup of individual files and programs. But they have a limited storage capacity and you must perform the backup yourself.

Backup to tape in an internal drive is also a possibility. You have more storage capacity and can set the drive to back up automatically at the end of the day. These tapes may contain only the changed data or all of it, depending upon how busy your operations are and how critical the data is to your financial health. You should back the entire system up at least once a week.

You can also back up to a remote backup service provider. The data is transmitted over phone, web, or optical lines and is stored at that remote location. There are some advantages to this; it isn't next door and you could have the data restored at a different location, should your own location be unavailable due to a disaster. This could speed up your recovery process, provided you have planned for an alternative operating location.

You can also back up to a remote location that replicates your data as described above.

No matter what type of backup method you choose, don't forget the software applications. You should have copies or originals of these stored somewhere safe and off-site. What some organizations do to get around the issue of out-of-the-box software is to have one standard setup of out-of-the-box software applications for every computer station in their institution. Those are the only programs they load onto each and every system. If your computer dies, then they just bring a new computer with the same software applications setup and you reinstall your data and away you go. If you load your own software applications, then you had better have a copy to install yourself because it won't be supported by your IT department.

### *Selecting What to Back Up*

You want to back up your critical data, your personal data, and any projects and databases you are working on. Think about how you would perform your job if your data was gone from your computer. If you cannot live without it, then it is critical. For example, the circulation system maintains a database of all books that are in circulation. There is a related database with all patrons' cards and personal data. If you restore the circulation system without the patron database, how will you check out books or get them back? Add to that the automated catalog and cataloging functions and see if the system is still operational. How does the loss of the catalog affect the circulation system?

**NOTES**

1. Message One and Blackberry are two of the companies that offer this service.
2. Frank J. Real, "Tick . . . Tick . . . Tick: Time Is Money When Recovering Lost Data," *Disaster Recovery* 15, no. 4 (fall 2002): 14–16.
3. Jason Buffington, "Data Replication Explained: Techno Advice for BC Pros," *Contingency Planning and Management* (May/June 2003): 58–60.
4. EMC Corporation, *A Symmetrix White Paper: Disaster Recovery as Business Continuity* (2002), www.emc.com/pdf/continuity/c894_disaster_recovery.pdf.
5. EMC, *Symmetrix White Paper.*
6. Peter G. Power, "Manage a Crisis, Don't Recover from Disaster," *Contingency Planning and Management Online* (January 2003): 22–26.

# 2

# Planning for the Worst

## *Loss of Computer Operations*

In the aftermath of September 11, 2001, the Federal Reserve, the Office of the Comptroller of the Currency, the Securities and Exchange Commission, and the New York State Banking Department met to draw up guidelines for the protection of data and the flow of information and monetary funds between U.S. financial institutions. Their concerns for computer systems were:

- rapid recovery and timely resumption of critical operations following a wide-scale, regional disruption;
- rapid recovery and timely resumption of critical operations following the loss or inaccessibility of staff in at least one major operating location; and
- a high level of confidence, through ongoing use or robust testing, that critical internal and external continuity arrangements are effective and compatible.[1]

In just the same way, libraries and archives need to be concerned with restoring access to their information and services for their patrons in a quick, efficient manner.

## Keeping Up with the Changes

### *Software*

Let's think about preventing the loss of data. What happens if you lose your data in a major server or network crash and it turns out that there isn't a current backup for the software. In fact, the software that you were using was the newest version of some program, and the computer center can't find that version and must use the previous one. Do you realize that you cannot open current data files with older software versions? You can do the reverse, of course: open old files with new programs. So here is another layer to add to backup routines and prevention programs: *back up the software.*

Of course, this disaster could be man-made if you "borrow" someone else's software, install it onto your computer (which is technically illegal unless it is freeware or shareware), and then your computer crashes and you find the only software you own is the older version. You are out of luck and lose all that time for data creation and access to your files, until you can purchase a new software program.

In the same breath, it is important to keep up with software changes. It isn't essential to purchase every upgrade, but pay attention to the industry. When it says that the new program no longer supports your software version, you have waited too long to upgrade and may have problems reading the old files with new software. Unfortunately the same rule applies to operating systems. If you wait too long to change to the new system, then you will find the data and its associated software won't run properly. In fact, this issue snowballs when you factor in the upgrading of your hardware.

When computers first became affordable, many cultural institutions delayed purchase until the "best" system could be found, only to realize that this was never going to happen. The institutions found that they had to buy into hardware and software, hope they had picked the right equipment and programs, and then get on