# Contents

**Foreword** ix  
**Preface** xi  
**Acknowledgments** xiii  
**Author** xv  

## Chapter 1 What Is Electronic Information, and Why Should You Care? 1  
1.1 Introduction 1  
1.2 Electronically Stored Information and the Federal Rules of Civil Procedure 1  
1.2.1 Changes to the Federal Rules of Civil Procedure 3  
1.2.1.1 Rule 1: Scope and Purpose 3  
1.2.1.2 Rule 16(b)(5) and (6): Pretrial Conferences; Scheduling Management 3  
1.2.1.3 Rule 26 4  
1.2.1.4 Rule 37 Safe Harbor 11  
1.2.1.5 Rule 34(b) Producing Documents Procedures 13  
1.2.1.6 Rule 33(d) Interrogatories to Parties 15  
1.2.1.7 Rule 45 Subpoena 15  
1.2.1.8 Form 35 15  
1.2.2 Federal Rules of Evidence 16  
1.2.2.1 FRE 502 17  
1.2.2.2 FRE 901 17  
1.2.2.3 FRE 802 18
# Contents

## Chapter 4  
**Who’s in Charge Here? Allies, Owners, and Stakeholders**  
4.1 Introduction  
4.2 The (Long) List of Stakeholders  
4.2.1 Information Technology Professionals  
4.2.2 Legal Staff  
4.2.3 Records Managers  
4.2.4 Auditors  
4.2.5 Human Resources  
4.2.6 Department Heads, Vice Presidents, and Executives  
4.2.7 Physical and Information Security Personnel  
4.3 Ownership of Data  
4.4 Data Control Considerations  
4.5 Required Skill Sets and Tools

## Chapter 5  
**The Hunt: Recovery and Acquisition**  
5.1 Introduction  
5.2 Where, Oh Where, Has My Data Gone?  
5.2.1 Applications as a Vital User Interface  
5.2.2 Hidden or Restricted Access Data  
5.2.3 Encrypted Data  
5.2.4 Deleted or Corrupted Data  
5.2.5 Proprietary Data or Data Stored on Obsolete Media  
5.3 Privileged, Sensitive, and Inaccessible Data Management  
5.4 Proving Ownership and Integrity  
5.5 Marking Time: How Time Is Recorded and Ensuring Integrity  
5.6 Legal and Forensically Sound Acquisition

## Chapter 6  
**Keeping Your Treasures: Preservation and Management**  
6.1 Introduction  
6.2 Securing the Data  
6.3 Access Control and Management  
6.4 Organization and File Management Techniques  
6.4.1 Day-to-Day Organization  
6.4.2 Management of Data over Time  
6.4.3 Response to Litigation or Audits  
6.5 Safe Storage Issues and Considerations  
6.6 Litigation Hold  
6.7 Spoliation: The Loss of Relevant Data  
6.8 Automated Technical Solutions

Click here to order "Electronically Stored Information, Second Edition"  
International Standard Book Number-13: 978-1-4987-3958-0 (Hardback)  
© 2016 by Taylor & Francis Group, LLC
## Contents

### Chapter 7  Sharing Is Good: Dissemination and Reporting  255

- 7.1  Introduction  255
- 7.2  Format Issues: Original or Usable?  255
- 7.3  Mediums for Transfer  259
- 7.4  Creating Readable Reports  261
- 7.5  Tips for Depositions and Expert Witness  264
- 7.6  Conclusion  266

### Appendix I: Links and References for More Information  267

### Appendix II: Forms and Guides  273

### Appendix III: Links to Technical Software Solutions  291

### Index  293
Foreword

Matthews has approached e-discovery from a fresh, new perspective—one that is understandable to the layperson as well as the technologist. Electronically Stored Information: The Complete Guide to Management, Understanding, Acquisition, Storage, Search, and Retrieval will guarantee that you know more about e-discovery than you thought possible. A must read for anyone in the information technology and legal professions, the book provides invaluable information to be proactive or reactive in responding to requests of electronically stored information. The flow of the book from the first chapter to the last is clear, simple, and thorough—any attorney who desires to become a technically savvy advocate for his or her corporate legal department or law firm will have this book at hand. This book goes a long way in removing the intimidation factor between IT, the corporate legal department, and outside counsel. This book should be required reading for anyone in a computer science, information technology, or law-related program, and is now part of the Digital Forensics and the Law course I instruct. If you want to get up to speed on e-discovery and actually understand what you read, you’ll buy this book.

Steve Hailey
President/CEO
CyberSecurity Institute
Digital Forensic Examiner and Educator
6

KEEPING YOUR TREASURES
Preservation and Management

6.1 Introduction

One of the most important tasks an organization or individual will face when dealing with electronic evidence is the preservation of that data in a way that ensures the integrity (guaranteeing that nothing about the data changes as you move and manage the data) and availability (ensuring the data are easily and efficiently accessible) of the data. In this chapter, we look at all of the aspects of that task and discuss the tools and requirements that you should be considering.

6.2 Securing the Data

Once you have gone through the process of acquiring electronic information for whatever purpose you might have, you will need to consider how you can secure that data (Figure 6.1). The first thing to consider is the actual classification of the data in question. Thinking about how you would classify the data you are trying to secure will inform you as to the degree to which you need to secure it. For instance, if the data you are working with are publicly available, easily recreated, and not subject to any question of legitimacy (either because the data contain well-established information or because they are simply not very important to the case at hand), you might classify the data as nonsensitive, low-security data.

On the other hand, if the data contain private information about an individual, protected information about intellectual property or critical infrastructure, or simply information that could be used by your adversary to your disadvantage, the data might be classified as
sensitive or confidential. Information considered work product or privileged in a legal definition would also be classified as confidential. Of course, if you are working for or with the military or an intelligence organization, there will be other levels of classification that determine exactly who has the right or ability to see that data. These can include, among others depending on the organization:

- For official use only (FOUO)
- Law enforcement sensitive (LES)
- Secret
- Top secret
- For your eyes only

Whatever the requirements of your organization or entity, the first step in deciding the level and processes of securing data is classification of that data.

Making decisions about how to secure information will always be a classical risk management exercise. You will weigh the risk in terms of what vulnerabilities or threats there are to the data storage, the likelihood of those vulnerabilities being exploited, the impact of that exploit if it should occur, and the cost of mitigating the risk to reduce the likelihood of an exploit resulting in the loss or compromise of the data.

In the first instance, when data are publicly available and well established as fact, there is little reason to protect the data. In risk management language, the vulnerability to loss or compromise will
be higher if the data are not protected, but the impact is minimal to none. Therefore, there is no motivation to spend resources protecting that data.

In a real-world example, if you have acquired information from an electronic version of *The New York Times* about specific features of a new version of smart phone that you will be using to establish some facts in your case, that information is publicly available and well established as fact. There will be other sources of the same information, again publicly available and easily accessible, which will corroborate your data. There is no reason or motivation for you or your organization to spend the time or resources to secure that information.

However, in the other instances you will not only have the motivation, but in many cases a legal and ethical responsibility to secure the data that you have acquired. Let’s look again at an example. If your company has intellectual property in the form of a patent for the technology that enables one of the features of that new smart phone and that could give competitors an advantage, the risk picture changes.

In this case, the vulnerabilities that might threaten the compromise or loss of your intellectual property (IP) might be the same as in our first example, but the potential impact is huge. Some of the possible threats to the loss of your IP might include

- Industrial espionage in the form of an attack on your computer systems
- An insider threat from a disgruntled employee
- An accidental loss of the information through the loss of a laptop, smart phone, or other digital device
- The compromise of your network or computing systems via a bad website or e-mail that installs malware
- The loss of data due to a natural or man-made disaster

The loss or compromise of that data could result in lost profits or business or might have a deleterious effect on a court case if you were arguing for your patent rights.

Here we see that your organizational risk analysis will indicate a greater need to expend resources and time as necessary to secure the data and mitigate the risks. Deciding whether to secure the data...
therefore becomes the first thing you need to consider before deciding how to do so. To accomplish that first step you will need to do the following:

- Take the time and expend the effort to understand the classification of your data and appropriately apply those classifications.
- Separate the different classes of data into unique groups for storage and/or management.
- Conduct realistic and thorough risk assessments of the data classes.
- Make decisions about the level of resources you will dedicate to the security of the data in their different classifications.

Once you have completed that first all-important risk management assessment, you are ready to consider the procedures and protocols for actually protecting the data.

6.3 Access Control and Management

The actual storage media for data is of relatively little consequence to the preservation considerations we address in this chapter. As long as the media is reliable, physically robust, and accessible, it does not really make a lot of difference if your data are stored on magnetic backup tapes, optical storage such as CDs or DVDs, solid-state drives such as USB devices, hard disks on a server, in a virtual server environment, or even on the Internet in some type of online infrastructure service (aka, the “cloud”).

When you consider the security and integrity of data, what really matters are the controls around who can access the data and how those controls are managed.

In this section, we look at various methods you should consider for managing access to your data. These can include

- Authentication
- Encryption
- Identity management
- Logical compartmentalization
- Physical separation and access control
We begin with authentication mechanisms as these are the most recognizable and common ways to control access to data (Figure 6.2). In nearly every organization, in order to gain access to your network, you use an authentication mechanism. In the world of IT we speak of factors of authentication. These include:

- Something you know—your user name or ID and a password, PIN, or a pass phrase that only you know
- Something you have—a smart card, token, or other identification device
- Something you are—biometrics such as your thumbprint, iris scan, facial characteristics, or even the speed and pressure with which you type on a keyboard

In most cases, only the single factor of “something you know,” consisting of the two pieces of information, your ID and password, is all that is required to gain access to your company’s network, your e-mail application, or even your home computer or smart phone (if you have bothered to enable that access control on your own devices).

However, when the classification of the data justifies a higher level of authentication, you will see two- or even three-factor authentication come into play. In those cases, you might carry a device around with you that you have to swipe through an installed reader on your computer, laptop, or other device before accessing information and
then follow up with your ID and password. Or the device might be designed to give you a onetime password or access number that you have to key in before putting in your regular ID and password. Or you might have to use the third level of authentication to prove who you are and that you have a right to access the data in question. You might have to scan your thumb, hand, face, or eyes or type with the correct and recognized patterns and pressures.

A more commonly used type of two-factor authentication that has become popular with some on-line applications (e.g., Gmail, Facebook, some banking sites) is the sending of a security code via text message to your phone before allowing you to login.

As noted above, all of these different types of authentication mechanisms will depend on the relative risk assessed for the data you are protecting. As you can imagine, the levels and costs of management and logistics for each of these types of controls increase substantially the more factors you introduce.

The next type of control to consider is encryption. Encryption can be the most secure way to protect your data. However, as with any type of control, there are trade-offs in efficiency and convenience. There are also many levels of encryption that can be used. A complete discussion of how encryption works and all of the different possible means to encrypt data is beyond the scope of this book, not to mention being beyond the mind capacity of most normal humans. Suffice it to say that you have choices if you decide that data are important enough to protect with encryption.

There are relatively simple means of encryption, such as password protecting a document or spreadsheet. This is accomplished simply by choosing the right menu item in your document software and typing in a password. The application will then encrypt the data in that document, so that anyone trying to view it without the password would not be able to open it in the first place. Even if they managed to open it with some type of plain text reader, it would be gibberish. However, this is relatively simple encryption and thereby relatively trivial to crack. Easily accessible tools that most anyone can find and install will be able to discover the password and allow someone to access that data.

Encryption types gain in complexity and therefore in levels of accessibility depending on your desires and needs. If you are protecting
state secrets, you can use encryption, the keys to which are huge and would take thousands of computers thousands of years to crack. And there are many levels in between.

Enabling encryption requires the management of those keys or passwords. In simple encryption models you only need to manage the passwords. But in more complex models there will be keys of different lengths and complexity that will need to be shared or protected in order to allow encrypted data to be shared and transported safely. Again, the complexities of key management are beyond our scope, but the point should be recognized: Encryption can be a logistical and management challenge.

New research into “quantum” encryption uses features of quantum mechanics and physics to create keys that will be impossible to recreate. Again, this is well beyond the scope of this book, but it presents new and very interesting possibilities for the protection of very sensitive data.

Encrypting data while in storage is only part of the process. If you need to transport or share that data with someone, you have to consider the ways and means of maintaining that encryption during transport.

If you are physically delivering the data, the data can remain encrypted on your storage device and you can provide the key to the recipient. If you are sending it via encrypted e-mail, you will need to have a method to share a key securely with the recipient.

If you intend to provide access to the data over the Internet or a network, you will need to provide the recipient with an encrypted “tunnel” (such as VPN, virtual private network) or an encrypted Web protocol (such as SSL, Secure Sockets Layer, or what you would see as an HTTPS web page; or SFTP, Secure File Transfer Protocol).

Again, the management, development, and logistics of creating and maintaining these controls exact a price in resources and time. The need to use these encryption controls will have to be weighed against the actual risk model that fits with the data you are protecting.

All of these systems and controls depend on being able to authenticate the identity of the person requesting access to data and to determine whether that person actually has the right to access the data. This is known as “identity management,” and there are a whole set of tools and protocols set up to facilitate this process. These tools
and protocols rely on databases of information that store identifying information and authentication credentials as well as the rights each identified entity has to certain stores of information.

Some of these identity management tools facilitate what is known as “single sign-on,” allowing a user to log in once to an organization and be authenticated via the identity management system to gain access to disparate sets of data or applications. This can increase efficiency and even security if the system is well designed and maintained. However, the adverse of that should be a caveat to anyone considering a single sign-on system. If not well designed and maintained, this type of identity management creates security issues that must be thought about carefully.

For instance, if you have a single sign-on system in place, your attack surface becomes less complex and now only depends on one failure to compromise all of the data that a user has access to.

That being said, single sign-on is often most appropriate when access is being granted to similarly classified sets of data. To allow single sign-on with access to nonsensitive data as well as highly confidential data would not be a good risk management decision.

Allowing access to your basic work files, e-mail, Internet connection, and desktop applications through a single sign-on would make sense, but access to a personnel database with confidential information might still require another level of authentication.

In the absence of this type of overall system, identity management is simply the connecting, via authentication credentials, of users or entities to the data or application to which they have rights. It will also manage different levels of rights such as read-only access (so you can view but not change data), write access (you can write new data), or change, delete, move, and so forth, rights that allow you to manipulate data depending on your authenticated permissions.

Another method of controlling access to data is to logically compartmentalize the data within a network or other storage medium. At the micro level, you can create profiles on one area of a computer or network drive or you can create partitions on a hard drive and assign different rights or access privileges to those profiles or partitions.

You can see an example of this on most any Windows-based computer by looking at the Documents and Settings folder. Within
that folder will be different names of the users on a computer. If there is only one user and if that user has administrative rights, then all of the subfolders will be accessible. However, if there are different users, they will only have rights to read or access or change the files in their own profile folder.

With different partitions on a drive, access can be created and managed through rights assignments or in some cases through the actual applications that control the startup or movement within the operating system. Different partitions can actually have different operating systems on them to which you must separately log in at the computer startup.

You can also use virtual machine software to create a computer within your computer. This type of software basically will allow you to set up a separate operating system with its own programs and data, all using your same hardware (such as the keyboard, mouse, monitor, DVD/CD drives, network connections, and USB or other connections), but existing as a file within your computer. Access to that system can be controlled via a password as well, and this can be used to store applications and data that are only accessible to those with the right authentication.

Virtual machines are becoming a common way to share space on servers in data centers and each of those virtual servers can have different authentication protocols for access to their data. This practice is becoming very popular because it saves resources in both physical space and in energy consumption by using one physical server device to host many different virtual server instances.

Within a network you can also create partitions or virtual networks. Similarly, these will only be accessible to those with the right credentials or sometimes even the necessary physical access. These can be divided using networking devices such as firewalls, switches, and routers that only allow certain traffic to access those segments of the network.

A final control that you can put in place with relatively little cost is simply to physically separate the data storage. This could be in the form of a limited access data center, locked server racks, or simply locked storage space where you keep tapes or hard drives.

As long as you maintain access control to those places and chain of custody documentation so you can document who has had access
and when, you will have a relatively reliable proof of the integrity and safety of the data in question.

6.4 Organization and File Management Techniques

To ensure the efficient availability of your electronic data, you need to consider the management of that data and how the data are best organized (Figure 6.3). In this section we consider different use cases that will suggest differing solutions to managing data. These include

- Simple day-to-day organization of data for business or personal use
- Management of information resources over time
- Organization and management of data in response to an audit or legal matter

6.4.1 Day-to-Day Organization

In the normal course of business or just your personal day-to-day life, organization of the ever-growing amounts of data that we create can be a daunting task. As we have noticed in our reviews of case law, courts and government entities such as tax collection agencies will require knowledge of your data and of its ownership and means of
access. In order to ensure that we have the requisite knowledge and access, there are several practices I recommend here.

First, as stated earlier, it is very important to classify your data. That classification process will be the first step in creating the logical and physical containers for your information.

Next, you or your organization should consider the different types of information you are creating and managing and make decisions about where and how the data should be stored.

This includes appropriate separation between an individual’s personal works while at the same time considering what data should be shared. To that end, you might create separated folders or network areas, some of which belong to individuals and are only accessible by them, while others are accessible only to certain groups of people.

There may be other storage areas that are accessible to everyone in the organization, such as an internal Web presence on which announcements are created and shared, or virtual bulletin boards or mailing lists. As we discussed previously, correct access to these different areas by the appropriate users will be managed by identity management and authentication tools.

One example of this might be a system of network drive letters that you see when you log in to your work computer. There might be one drive letter that accesses only your personal information. Only you and the system administrators will be able to access or change information on that drive or folder.

There might be a different drive letter pointing to a network storage area on which your organizational division or department shares data specific to your tasks. Finally, there might be drive letters that point to storage for organization-wide data, or possibly administrative or system data.

Managing the data within each of those storage areas is equally important. As an organization, it can be extremely valuable to take the time to create a logical file system that meets and reflects the business protocols and strategies. As an individual, this can be an equally valuable exercise. In most cases, this is rarely done sufficiently if at all. And in those cases, it is often an afterthought or something that you are forced to do at some point in order to find important or relevant data, or an audit or other regulatory compliance issue might require this type of organization.
Taking the time to consider the best way to organize your data is a worthwhile exercise whether it is at the beginning of setting up a system or after it has already been developed ad hoc.

Start by thinking about the good old-fashioned file cabinet. Many of you probably still use these for storage of paper documents. I suggest having the business managers for your organization (or possibly just your division if yours is a large organization) get together and do a little brainstorming about what are the important aspects of your business. You can also do a similar exercise on a personal level.

Outline the different important things you will need to keep track of and the subcategories underneath them and any categories that might fit beneath those. Consider the example of a small, family-owned gift shop. Some of the aspects of their business will include

- Facilities
  - Rental or mortgage payments
- Insurance
  - Liability
  - Inventory loss
  - Structural loss or damage
- Maintenance and janitorial
- Inventory
  - Purchasing
  - Supply chain
  - Shipping
  - Resupply and auditing
- Employees
  - Compensation
  - Insurance
  - Taxes—withholding
  - Health care
  - Vacation and sick leave
  - Other benefits
- Management
  - Compensation
  - Benefits
- Taxes and fees
• Local
  – Business licensing
  – Fees and dates of renewal
  – Other taxes, dates due, audits
• State
  – Licensing
  – Taxes
  – Other fees
• Federal taxes
• Marketing
  • Advertising
    – Internet
    – Magazines, newspapers
    – Yellow pages
• Web presence
  – Company website
  – Social media
  – Specialized gift-shop sites
• Financial
  • Auditing
  • Accounting
  • Procurement
  • Payment methods
    – Credit card online or point of sale
    – Cash
    – Checks

As becomes obvious in this exercise, even a very small organization has a great deal of complexity and areas of concern that are best considered and managed carefully beforehand.

However, there is no reason such an organizational exercise cannot take place later in the process as well. It will always be valuable.

Once you create this outline, consider how you would organize those files in a file cabinet.

You might have different drawers for each year of your organization’s life as a way to start organizing. Then you would take each of the above major categories and create a divider alphabetically within the drawer for the current year.
Within each of those dividers could be hanging file containers labeled with the subcategories. Finally, within those containers could be file folders containing the actual documentation of the sub-subcategories.

Of course, the blessing and curse of the electronic version of this system is that you can keep going deeper and deeper with sub-sub-sub-subcategories ad infinitum. That is harder to do in a physical file cabinet as you soon run out of room to stuff more folders.

The blessing can be that you are able to more precisely and specifically organize all of the intricate and complex parts of a business system. However, it is quite often the case that these categories and sub-sub-subfolders are not well considered and become redundant, unnecessary, and extremely difficult to manage.

If you are able to carefully consider the actual needs of your organization and carefully manage the way the data are stored, you could take our example above and create a logical electronic file system in which each aspect of your business is logically segregated and easily managed. As long as it is a part of your accepted business practices to maintain a regularly monitored and carefully managed file system such as is outlined in our example, you are likely to have achieved the goal of efficient and easily accessible data.

If that has not been the case, it is never too late to spend the time considering how things should look and reorganizing your data into the properly labeled and segregated file system.

In one organization I worked with, we had a quarterly meeting wherein we would gather and look through our file systems together, clean out the redundant or unnecessary files, and move or reorganize files as appropriate. It was often the case that we would find a series of files or folders labeled with a similar date or title and realize they could all be consolidated into a single folder with that name. That exercise was both a great team builder and a wonderful way to better manage our file systems. I highly recommend it.

6.4.2 Management of Data over Time

When we consider the management of data over time, we are most often concerned with records retention rules and regulations. Depending on the organization you work with, or your personal
business practices, there will be differing rules that apply to you for how long you are required to store electronic information. Each state and some local governments will also have specific records retention laws that relate to records held by government entities. The federal government also has records regulations specific to government and to some types of businesses such as the financial sector.

Actually outlining all of the specific rules, laws, and regulations would require an encyclopedia collection, and by the time you finished you would have to start over again because it would have all changed. So, we will not even try to do so here. However, it is a very important business requirement for any organization, or even for you as an individual, to take the time to understand those rules and develop a system to ensure you are complying with them.

For instance, in the local government organization for which I once worked, we had state-mandated records retention laws that stated the specific number of years we were required to keep documents such as contracts for purchases, legislative analyses or decisions, financial records, network logs, and so forth. Our records management team worked diligently and endlessly

- Interpreting the different laws and regulations that applied to each part of our organization
- Training staff and management on those requirements
- Assisting departments with records management systems and designs
- Consulting on the development of organization-wide systems and tools for the efficient management of data

Through their efforts we created e-mail–based archival tools that allowed us to select which e-mail and attachments needed to be saved for a specific amount of time based on their classification and the retention rules that applied.

We also had systems in place for storing and archiving other documents and electronic records based on their designated records retention categories. We were not a huge organization, but this was still a pretty difficult challenge at times. It took a long time and a great deal of effort on their part and a commitment on the part of our management and executives for us to be able to begin to get a handle on these requirements. It was and is made harder by the general
tendency toward entropy and disorganization that is inherent in a large, dynamic, and diverse organization.

I want to be clear that I am by no means implying this is a simple or trivial exercise to complete. But I will state with confidence that to whatever degree you can manage your file systems up front or as you continue your business or personal management processes, you will be greatly enhancing your ability to comply with records retention requirements. This may only become important when the auditor or tax authorities come to visit, but it behooves all of us to be ready for that before it happens.

6.4.3 Response to Litigation or Audits

When we are considering the organization and management of our electronic data, one of the most compelling reasons for careful and consistent management is the inevitable need to produce those records for a legal action or audit. Whether you are working with an organization or simply thinking about your own records, the chances are very high that you will at some point be responsible for locating and acquiring those records and managing them appropriately for production to a court or auditor.

All of the methodology we outlined above is the basic foundation of this process. We will discuss in more detail in the following pages the different types of systems that you may need to put in place.

First, though, we look briefly at the different parts of a good production process. If you think about the whole of the data that either you or your organization has accumulated as a large pie chart, in the end there is a very small subset of that whole pie chart that will be relevant and required for litigation or an audit. To get to that slice of the pie you will need to go through several steps.

First is acquisition of the data. To do this efficiently will depend a great deal on your having done your due diligence and created a well-designed and managed file system. You will need a good understanding of all of the places that relevant electronic data exist as we discussed in great detail in Chapter 3. It will be equally important to be able to locate the people who own and control the data as we discussed in Chapter 4.

When you have located all of the relevant data and contacted the custodians to gain access to that data, you will need to carefully acquire
the data in a forensically sound manner that preserves its integrity as we discussed in Chapter 5. Now you understand better how all of this works together (and why you need to read all of those chapters!).

Acquisition is the first part of the production process and should result in a smaller part of the pie. In this first step, you will be painting with a broad brush and acquiring anything that might possibly be relevant. As we saw in our look at case law and the rules of court procedures in Chapter 1, the courts require that you preserve anything and everything that might be relevant.

If you have done a good job of organizing and have followed your records retention rules, you should have a pretty good handle on what that subset will consist of and how to gather and preserve it.

An important point that we made earlier needs to be restated here. Your ability to preserve the required data is often predetermined by your or your organization’s actual records management practices. A court or auditor will be looking at those practices and not at any written policies or procedures.

The fact that you have well-documented records management rules means nothing if your employees either do not know about them or have not been following them. A great example of this is a case where an organization claimed that their e-mail destruction policies were the reason that certain important and relevant e-mails were deleted before they could be preserved. In deposition, an administrator testified that the organization had never followed those destruction policies until the day after they were served with notice of litigation. Needless to say, that did not fly well with the court, and the organization (and their counsel) was slapped with heavy sanctions for spoliation of evidence.

Once you have carefully acquired and preserved the evidence, the next step will be the review of the evidence. This is often done by paralegals if it is a legal matter. Their job is to sift through all of the preserved data with the facts of the case in mind and winnow out the actual relevant evidence. One way this can be done is by searching through the preserved evidence for keywords. Often the people tasked with reviewing and producing the relevant data will begin by establishing a set of words or phrases related to the case at hand. Automated systems or applications or even command lines can be used to search through large sets of electronic data for “hits”
on these keywords. Those hits offer the worker a first hint as to which of the data in question might actually contain relevant information. However, keyword searches do not always do a thorough job of finding all relevant data. Words in documents might have been spelled incorrectly or somehow corrupted so that those relevant pieces of information are missed. Or the keywords you are using might simply not be sufficient to find all of the important data.

Keyword searches can be an important first step, but they are not the be-all and end-all that some organizations consider them to be, and courts and auditors may demand a more thorough review of preserved electronic information to find all possibly relevant evidence.

In fact, in a recent case opinion by Magistrate Judge Andrew Peck of the South District of New York, he approves the use of “computer-assisted” review of evidence and discusses the benefits that can be derived from these relatively new artificial intelligence technologies.* These tools allow legal professionals to train the application to search for and discover relevant electronic evidence much more efficiently than a simple keyword search.

The searchers will also need to look for and categorize any evidence that would be considered privileged or attorney work product and remove that from the producible information.

Once all of that has been accomplished, the paralegals or attorneys will organize the producible data into categories and groups based on the facts and issues in the case, or possibly any agreed-upon production criteria of data for the opposing party.

For instance, the parties may have agreed to produce e-mail records for certain date ranges and regarding certain subjects or keywords. They may also have agreed that those e-mail records could be produced as TIFF files or images with or without copies of the metadata that revealed the source, destination, path, and date/time stamps of the e-mails. They might have required that they be produced in their original format which is quite often the case as that can do a better job of preserving metadata. In this instance, the legal team, after

---

removing non-relevant or privileged e-mails, would preserve the rest of the e-mail data in the fashion that was agreed upon.

At this point we have reached that small section of the pie that is producible, relevant electronic evidence. As you might imagine, this process can be extremely time consuming and resource intensive. But the degree to which your organization has developed and managed its records will make a huge difference in how onerous or simple this process is for you.

6.5 Safe Storage Issues and Considerations

Keeping your data physically and logically safe is one important component of establishing and ensuring the integrity of the data.

In Section 6.3, where we discussed authentication, we discussed the idea of securing your data logically by virtue of good identity management systems, encryption, and segregation of different systems. Those means of securing the access to your data are your first line of defense in ensuring the safe storage of your information.

A second consideration is the actual physical storage and your disaster recovery and business continuity planning. These should include backup and restore procedures as well as emergency readiness and resilience considerations (Figure 6.4).

The physical space where you store your data will need to be secured with access controls. If you are an individual, that could be as simple as locking your laptop or desktop to your desk with a cable lock. A small organization might simply have a locking closet or small

Figure 6.4 Having a disaster recovery plan is important.
room where you store your servers and other networking equipment. However, keep in mind that part of your secure storage responsibility will be ensuring that the devices in those storage areas are not subject to overheating or power outages.

You will need to ensure that there is sufficient cooling capacity and some type of redundant power supply or battery backup for those storage devices. Computer systems can be destroyed by overheating, and if they shut down unexpectedly due to a power outage, data can be destroyed.

In a larger organization there will most often be a special data center. It will have robust physical security including guards, cameras, fences, and often several layers of secured access using authentication mechanisms such as card or fingerprint or other biological scanners, all of which are carefully monitored and recorded.

There will also be redundant systems for cooling, heating, and power backup systems. If you have never visited a large data center, I recommend arranging a tour. They can be quite impressive and complex. One very large center that I toured had multiple large diesel generators and huge reservoirs of diesel fuel. They also had a system that recycles all of the heat generated by the servers to heat the building, and they have managed to cut their energy consumption way down with some amazing and efficient systems.

Whatever your physical storage medium, you should ensure it is safe from intruders and disasters to whatever extent is possible and makes sense from a risk management perspective.

It will also be important to test those systems on a regular basis to ensure that the redundant cooling and power procedures work in an emergency and to consider how you would respond if they failed.

You also need to consider a robust and consistent backup procedure whether you are an organization or simply dealing with your personal data. There are many different backup systems available depending on your specific needs. For a personal system, you can either back up your data on a regular basis to an online storage system or back up to a connected device such as an external USB drive or a separate part of your hard drive.

Online or cloud services are becoming quite ubiquitous and can be relatively inexpensive or even free depending on the amount of data you need to store there. You need to be cognizant of some caveats,
however, if this is your choice of backup medium. When choosing to store your data on an Internet site, you are basically extending your network or data set to a physical and logical location that is no longer in your immediate control. In fact, by the very nature of Internet or cloud storage, your data could be residing on a server in another country thousands of miles away. It could be administered by poorly paid and poorly vetted employees of an outsourced contractor for the organization to which you thought you were committing your data. The organization might even be a couple of guys in a garage with a small server center. They might have little or no physical or logical security in place, no backups, and no recovery plan in case their garage catches fire or floods or some authorities come in and confiscate their (your) servers.

On the other hand, the organization might be very well established, only hire employees who have been extensively background-checked, regularly back up their (your!) data, and have well-established and regularly exercised security and recovery plans and procedures in place.

The point is that unless you are very cautious and do your homework, storing your data on an Internet-based service can bring a whole new set of risks to consider.

All of the regulations and requirements we have discussed as your responsibilities for the management and production of your data are still yours even though you have contracted with another organization for the storage and backup of your information. As such, it is inherently your responsibility to ensure you know how your data are being protected and secured as well as how you can gain access to your data at any time.

The best way to do this with an online or cloud service provider is to ensure that you have strong and comprehensive contract language that covers all of the issues we outlined above. Make sure that you have a signed agreement with the provider that you both agree to regarding

- Day-to-day security of your data—ask for a copy of their security policy
- Vetting of their employees
- Backup procedures and timing
• Patch and configuration management of their infrastructure
• Disaster recovery and business continuity plans for recovery and reestablishment of your data
• Notification procedures if any data are compromised, destroyed, or lost for any reason
• Up-time guarantees—should be as close to 100% as possible
• Support procedures and contact numbers, including hours of support and backup support options
• e-Discovery or data production procedures—how do you access and acquire your data if the data are needed, in an efficacious manner

The most important thing you can do once you have such a contract in place and feel satisfied that all of your concerns are addressed will be to test! Ask your provider to demonstrate that their security can stand up to a third-party penetration test or that they have an auditor buy off such as a SAS 70 certification.* Ask them to do an exercise with you where they and/or you have had a disaster or some other need to recover your data and ensure that those procedures work as advertised. A good contract agreement is the first step and an important one, but testing to ensure all of the processes actually work is equally important and should be done on a regular basis.

You can also choose to back up your data via a physical device such as a USB external drive or onto CDs or DVDs or onto backup tape systems. This can be a good choice, but again it will require you to have established, well-documented, and carefully followed procedures in place.

You should have a scheduled and if possible automated backup date and time. It should include full backups on a regular basis and incremental backups (where you only back up data that has changed or been added) more frequently.

Consider where the backed-up data will be stored and whether you will want to have more than one copy. This is again a risk management decision that should be based on the classification and importance of the data.

---

It does not make sense to spend the time and money to create an extensive tape library, keep multiple redundant copies, and have both on- and off-site storage if the data you are storing are readily available and nonsensitive information. On the other hand, if the data are proprietary or confidential and difficult or impossible to re-create, you might be justified in expending the time, money, and resources to ensure the data are extremely well protected.

Finally, an often overlooked part of backup systems is the actual restoration procedure. Creating and carefully storing backups of your important data is all very well, but if you do not have or have never tested your restoration procedures, those backups are worth very little.

Backup systems can be complex and subject to failure. Maybe you are an organization whose disaster recovery plan relies on restoring all of your critical data from off-site backup tapes, or simply an individual who wants to be able to rebuild your computer if you have a hard drive crash. In either case, taking the time to test your restoration procedures on a regular basis is extremely important.

If you are a large organization, you should first ensure that you have a disaster recovery/business continuity plan in place and that all of your employees and managers are aware of it. Then you should schedule regular exercises to test that the procedures outlined in the plan actually work in real life.

For instance, do a simulated live-fire test of your processes. Make sure that you have the contact information for your off-site storage provider and your data center recovery staff members. Find out what it would take in time, money, and resources to actually re-create your business after a flood or earthquake or bombing event.

As an individual, take the time to regularly test the recovery process of your backup system. Usually that simply means firing up the backup application and selecting “restore.” It is a great idea to try this just to understand what it will take and to ensure that your backups have not been corrupted.

I also recommend off-site storage for both individuals and organizations. If you are a victim of physical theft or damage, on-site storage of your backup data might suffer that same damage or be stolen along with everything else.

Organizations should contract for secure off-site data storage. Individuals can simply take their external hard drives or CD/DVD
backups and store them away from their premises, somewhere safe and easily accessible.

The bottom line is that backups and off-line storage can be a final and extremely important resource for controlling access to your electronically stored information. Think carefully about the most effective way you or your organization can manage these procedures and ensure they are done correctly, practiced religiously by everyone involved, and tested regularly.

6.6 Litigation Hold

One important process to be aware of and ensure is well established, especially in an organization, is the litigation hold.

In any legal matter in the United States, you are required (as we illustrated in Chapter 1 on rules of procedure and case law) to carefully preserve any and all relevant evidence as soon as you have reason to believe there might be litigation (Figure 6.5). That is a relatively

![Image of a person holding a clipboard]

Figure 6.5  Creating a good litigation hold process will help protect your organization from liability.
vague and anomalous definition of when you should actually trigger that preservation process, and it is subject to interpretation. However, in most of the cases I have studied, it was clear that courts expect you to begin preserving data as soon as you would reasonably expect that there might be a legal matter arising from your actions or from some event.

There have been cases where the courts ruled that data should have been preserved because someone received a phone call from a known litigious company regarding their patent rights. They did not specifically say “we plan to sue you” or serve any kind of legal notice. However, the courts basically said, “You should have known this would become a legal matter and started preserving data immediately—and since you didn’t, you are liable for the loss of that data.”

As an organization, you should spend time with your legal counsel discussing your policy about what constitutes a trigger for preserving evidence. Again, as always, this is a risk management decision about how much time, money, and resources you feel are appropriate to spend on the preservation, management, and secure storage of data that might possibly become relevant electronic evidence.

Once you have established and documented your decisions around what should trigger that preservation, you should create a litigation hold policy and procedures. This policy should include the following details:

- Who will be the responsible party for notifying any and all owners of relevant data that they are required to preserve and produce that data (this will usually fall to the legal counsel)
- A process for establishing where all of the relevant data reside and all of the different sources and devices that might hold that data
- A record of who the owners of that data are and their contact information
- A process for ensuring the security and integrity of the data
- Who will be responsible for the acquisition and secure storage of the data
- A system of auditing and verifying that all relevant data were preserved (in my former organization we had data custodians sign an affidavit that they had searched for and produced any
and all relevant evidence that they were aware of and that was in their control
• An ongoing monitoring and maintenance of the hold and preservation of any new data

In my former organization, our law department was the responsible party (and most case law puts the responsibility for litigation holds squarely on the backs of legal counsel). At the trigger event, whether that be an employee being let go or threatening some action, or some event that creates liability, or any other event that they consider a trigger, the attorney or their paralegal would set up a litigation hold meeting. In that meeting we invited the owners of the data, their managers, their IT staff members, and the legal personnel to all get together and discuss all of the places that relevant information might reside.

The legal folks outlined the parameters such as dates and subjects that were in scope. The IT staff was there to assist with understanding the logistics and procedures necessary to acquire, store, and secure the information. We also sometimes included an electronic discovery expert to assist with ensuring that all possible areas of evidence had been considered. All custodians of the data were given affidavits that they were to sign and return to the legal staff when they believed they had preserved all relevant data that they controlled.

After that initial meeting, the legal department was responsible for sending out formal notifications to all affected employees and managers that outlined exactly what they were responsible for preserving, how they should preserve it, and for how long. They were also responsible for following up and ensuring that any new data were preserved and all data were kept until no longer needed.

This type of system and policy will go a long way in complying with what the courts and auditors expect of you and your organization. It will be of value to you to create such a system and put it to use when appropriate.

6.7 Spoliation: The Loss of Relevant Data

We discussed spoliation in Chapter 1 when we looked at the rules of civil procedure and other rules that apply to litigation. An example
of spoliation is best shown in *West v. The Goodyear Tire and Rubber Company* in 1999, which found that spoliation is “the destruction or significant alteration of evidence, or the failure to preserve property for another’s use as evidence in pending or reasonably foreseeable litigation” (Figure 6.6).*

As we consider this in light of the subject of this chapter, keeping our data safe and preserving its integrity, it should be obvious that the procedures and policies we discussed earlier are the most important ways that you as an individual or an organization can avoid spoliation of relevant evidence.

Spoliation sanctions can be administered in cases whether or not the loss of data was inadvertent. If you have good data retention policies and practices in place and can prove to a court or jury that you have followed those policies, it will go a long way toward limiting your liability for lost relevant electronic evidence.

On the other hand, if you have not put those policies and procedures in place or have failed to practice them consistently, the courts are much more likely to administer more severe sanctions for the spoliation of evidence.

Sanctions for spoliation are common in litigation involving electronic evidence. This is because these types of cases often involve large amounts of complex information and because many organizations and individuals have not made the effort to understand all of the issues and ramifications of our new world of electronic data.

By virtue of your having taken the time to read this book and take other steps to both understand and begin to manage your electronic information, you gain a great advantage. If you carefully create the foundation of a good records retention policy, understand and manage your electronic data inventory, and create and maintain a well-considered and monitored litigation hold process, you will have a much better handle on where your data live, who owns that data, and how to retrieve and preserve that data and the integrity of the data. The chances that you might inadvertently lose or compromise electronic evidence that is relevant to a case or an audit will decrease immensely.

Being aware of and preparing to avoid spoliation will save you and your organization money, time, and resources. The threat of spoliation sanctions is simply another good reason to take the time and effort to develop the policies, procedures, and practices outlined in this chapter.

6.8 Automated Technical Solutions

In this final section, we discuss some of the ways that automated solutions have been designed to assist in the procedures we suggested above.

It is not the place of this book to specifically advocate for one solution over another. And in fact, as with anything involving technology, these solutions are forever evolving, and any that we might recommend would very likely be out of date in the very near future. We will, however, list some of the current solutions that are available in Appendix III (Figure 6.7). Here I will give you some guidelines to what some of the best solutions can provide and what you or your organization should be looking for if you decide to either purchase or develop such a solution.

The primary consideration for any solution is whether it can address all of the goals we discussed in this chapter, such as:
Keeping Your Treasures

• Ensuring the integrity and security of the data
• Creating a methodology to expedite and enhance your ability to locate data
• Assigning and managing ownership and authentication
• Assisting with the secure storage and preservation of electronic evidence
• Managing notification messaging systems
• Enhancing your ability to analyze and categorize data

When you make the decision, either as an individual or as an organization, to develop the capacity to manage your data and be prepared for whatever need might arise for the production of that information, I offer the following checklist as guidance:

• List and engage your stakeholders: Consider who should be involved in the decision making regarding records retention, litigation holds, and so forth. This might include any or all of this list.
  • Legal counsel and paralegals
  • Auditors
  • Management
  • IT staff and managers
  • Information security professionals
  • Database administrators
  • E-mail administrators
  • Human resources
  • Union representatives
• Executives
• Risk managers
• Records management staff and archivists

• Create a business justification: What are the goals and business reasons for either purchasing or creating this solution? Document the specifications of the solution that will meet those goals, listing the benefits and the risk assessment and mitigation decisions that justify each of those specifications. This important document will both guide your decisions on the best solution and assist with justifying your decisions to business leaders. Some of the possible specifications might be

• Data-mapping tools that mitigate the risk of compromise of data that could result in data loss, spoliation sanctions, or loss of data integrity
• Secure transport and storage capabilities that address the risk of a loss of data integrity or compromising of the evidence
• Authentication and identity management tools that mitigate the risk of mishandling of data or access by nonauthorized personnel
• Cataloging and indexing features that enhance and expedite your ability to categorize and classify data, thus addressing the risks of inadequately protecting sensitive data, or adversely, spending unnecessary resources to protect publicly available data (this can also mitigate the risk of your producing work product or privileged evidence)
• Notification and messaging systems that mitigate risks of having failed to adequately protect data through notices to owners and custodians of relevant electronic data, and the possible threat of spoliation sanctions that might result

• Research possible solutions or combinations of solutions that will best meet your compiled list of specifications. These might include

• Off-the-shelf solutions, some of which we reference in Appendix III
• Something you are capable of developing in-house
• A hybrid combination of in-house solutions and off-the-shelf tools
• A tool or system that is physically or logically run from within your organization
• A service located on the Internet or hosted by a contractor
• Assess all of the possible solutions with your stakeholder team, considering
  • Which solution or combination of solutions best addresses your specifications (some specifications might have greater weight than others—these decisions should be made up front by your stakeholder team)?
  • Which gives the most value for its actual costs (do not forget to include training, setup, management, and ongoing operations and maintenance)?
  • Which is the most feasible for deployment in a reasonable time frame?
  • Which will be scalable, robust, and resilient enough to serve you or your organization into the future should you evolve or grow, or when the technology evolves as it is sure to do?
• Finally, document your decision and ensure you have buy-off from all stakeholders, especially the executives and leaders who will be paying the bills.

The purchase, development, and deployment of an automated solution can be of great value in avoiding all of the risks we outlined in this chapter and enhancing your ability to address all of the requirements for keeping your data safe.

It is absolutely worth your time to consider what options are available to automate the procedures necessary to provide for the integrity and security of your electronic information.