Addressing Today’s eDiscovery and GRC Challenges

Governance, Risk, and Compliance regulations surrounding discovery of electronically stored information have changed corporate business practices. With expanding rules and regulations, the legal department must defend the corporate interest, meet governance challenges, and ensure regulatory compliance, while balancing risks and expense. To meet today’s challenges, corporate attorneys need legal knowledge and experience, as well as an increased understanding of corporate IT capabilities, policies, and technologies. At the same time, corporate IT needs to understand the significant risk and cost implications of implementing technologies that do not address legal, compliance, or regulatory requirements.

The authors of this book are a part of the EMC Information Governance, Risk, and Compliance Practice, which is a specialty practice within EMC Corporation, consisting of seasoned attorneys who also have backgrounds in business law, eDiscovery, information technology, records management, security, and compliance. The team helps organizations understand their legal and regulatory obligations, best practices to meet those obligations, and how EMC solutions can be leveraged as a part of a GRC or information governance strategy and process.

For more information about the EMC capabilities and solutions, visit www.emc.com/informationgovernance or contact us at goodgovernance@emc.com.
The Lawyer’s Guide to eDiscovery Technology for Dummies

The world creates and stores data at incredible rates. In 2009, the Digital Universe — all of the world’s digital information — grew to 800,000 petabytes, equivalent to a stack of DVDs reaching from the earth to the moon and back, or one million times the amount of information in the Library of Congress. The IDC Corporation forecasts growth reaching 1.2 million petabytes, or 1.2 zettabytes by the end of 2010, and then to 1.8 million zettabytes by 2011.

A petabyte is a thousand terabytes or a million gigabytes.

The bad news for companies like yours is that IDC also says that enterprises are liable for about 80 percent of that data. It’s a huge task made even more complicated because of the amazing web of compliance and privacy laws and requirements that require chunks of that data to be retained, deleted, stored, or secured at different rates.

Even if you’re not particularly tech-oriented, you’re a lawyer. That means having to deal with both that stored data and the laws that concern it. You work with companies to locate and preserve data that is relevant to a specific lawsuit, which can be a daunting task if you think about the massive amount of information and all the different repositories that may hold that information.

If you do this work poorly, you and your company can face serious court sanctions, ranging from small money penalties to outright legal judgments being entered. The bad press that comes along with the sanction can be even worse!

Because of this, many companies implement initiatives to improve their eDiscovery processes. Even then, it can be a difficult process that requires a lot of communication and coordination between two departments that just don’t seem to speak the same language: Legal and IT.

In fact, without personnel who can understand both legal and IT issues, initiatives frequently falter or fail.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
About This Book

This book is written to help personnel in both the legal and IT departments better understand those parts of the eDiscovery process that may not come naturally to them. Hopefully, what you read here will bring the two sides together in blissful eDiscovery harmony.

The side you’re reading right now assists lawyers in understanding more of the IT-specific requirements of the eDiscovery process. The other side of the book will help people with strong IT skills to better understand the legal parts of the eDiscovery process. On the off-chance that these two groups are fighting over a copy of this book, please either tear it in half or call EMC for another complimentary copy.

Icons Used in This Book

This book uses a couple of icons to draw your attention to special information. You’ll gain some good knowledge if you stop and take notice.

This icon denotes information that will save you time or money.

This icon points out common traps and pitfalls you might run into. Pay attention to these bad boys — it’s better to be warned now than run into trouble later.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Chapter 1

Preparing for and Initiating eDiscovery

In This Chapter
▶ Learning what eDiscovery means
▶ Examining the Electronic Discovery Reference Model
▶ Managing your information prior to eDiscovery

eDiscovery is necessarily a reactive process, as you only perform eDiscovery in response to certain events. That shouldn’t stop you from preparing in advance and reaping the benefits of your foresight. In this chapter, we look at the various forms of eDiscovery and how companies can prepare for the eDiscovery process.

Defining eDiscovery

The term eDiscovery normally conjures the specter of litigation for most lawyers. But in today’s compliance environment, the process of eDiscovery takes many forms. When you define eDiscovery as the process of identifying, collecting, and evaluating information as part of a legal inquiry, it becomes very similar to several other activities. Take a look at Table 1-1 to see those similarities.
Table 1-1  eDiscovery Similarities

<table>
<thead>
<tr>
<th>Process</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Litigation</td>
<td>Providing ESI as part of a lawsuit, whether you’re a plaintiff or a defendant.</td>
</tr>
<tr>
<td>Regulatory inquiries</td>
<td>Providing ESI to a regulator investigating you or checking on compliance requirements.</td>
</tr>
<tr>
<td>Internal audit</td>
<td>Implementing internal audit processes to ensure that certain policies and procedures are being followed — because most information is now electronic, these auditors must request or find ESI on their own.</td>
</tr>
<tr>
<td>Freedom of Information Act/Open Records Requests</td>
<td>Processing requests from citizens for government-held information.</td>
</tr>
<tr>
<td>Third-party subpoenas</td>
<td>Processing requests for information for entities not involved in litigation as a named party. Although non-party obligations are usually less onerous and costs can often be recouped, the process of finding and producing this information is virtually identical to the litigation process.</td>
</tr>
</tbody>
</table>

Although these processes have a lot in common, the starting points can vary greatly for each, as can the time available to respond and the penalty for failing to get it right.

The eDiscovery Process

Think of the eDiscovery process as an enormous funnel. You shovel electronically stored information, or ESI, into the top of the funnel and run it through a series of filtering processes that cut down the amount of information (mostly to eliminate duplicates and irrelevant or privileged data). Once all the filters have done their jobs, the final data set comes out as the product of the process.
Chapter 1: Preparing for and Initiating eDiscovery

The keys to an efficient process are:

✓ Restricting the amount of information shoveled into the top of the filter, because it’s expensive to push data through the process.
✓ Making sure that potentially relevant data is fed into the filter — if you miss relevant data, you may face sanctions.
✓ Making the filtering process as fast and inexpensive as possible.

Understand your current legal and regulatory profile as the starting point. You need to know what type of lawsuits are common, what data is involved in those suits, which regulators actively seek data, and any other relevant information.

Sounds simple, right?

Introducing the EDRM

eDiscovery is a complex and ever-changing process, so it’s helpful to have a useful framework with which to analyze the process. Take a look at the Electronic Discovery Reference Model, or EDRM, in Figure 1-1.

![Figure 1-1: The Electronic Discovery Reference Model.](image)

Although the EDRM process moves from left to right, you’ll note arrows going in every direction. The process can loop back and forth depending on new information about the case, adding or reducing ESI, and changing legal claims and theories.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
We use the EDRM model throughout this book; the model itself can be broken down into three basic parts.

**Information management**

This node refers to things that happen “outside” of the eDiscovery process, but still have a significant impact on the process itself. For more on this see the section “Understanding Information Management” later in this chapter.

**Identification, preservation and collection**

The left side of the model includes steps that are focused on IT issues: identification, preservation, and collection. ESI within the infrastructure that is relevant to a case must be identified; then it must be preserved (so that it’s not deleted) and eventually collected (for further work).

**Processing, review, analysis and production**

The right side of the model includes tasks that are more legal in nature: processing, review, analysis and production. In these steps, ESI is processed so that it can be reviewed and analyzed for its impact on a case. Ultimately, relevant, non-privileged information must be produced to the other side in the case. (The last step, presentation, refers to how data is presented at trial, which is rare).

**When Do We Start?**

The duty to preserve data actually can begin long before a party is sued or files suit. The Sedona Conference Commentary on Legal Holds says

Reasonable anticipation of litigation arises when an organization is on notice of a credible threat it will

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Chapter 1: Preparing for and Initiating eDiscovery

become involved in litigation or anticipates taking action to initiate litigation. In other words the threat must be a credible probability — and not just possibility.

Waiting to preserve ESI until after receiving a summons and complaint could result in spoliation claims and sanctions.

When you can reasonably anticipate litigation, you must take some critical steps quickly:

1. **Suspend any records management or other policy that allows the destruction of potentially relevant information (both ESI and physical data).**

2. **Send a written litigation hold notice to all custodians of data that may have discoverable information.**

Don’t rely on the notice and the custodians as the only method of preserving ESI. Depending completely on custodians to preserve relevant ESI has been heavily criticized in many recent cases.

3. **Follow up with custodians that have questions or concerns.** You have a duty and an obligation to stay current and involved with the process.

4. **Ask each custodian to acknowledge receiving the notice.** Send reminders on a regular basis.

5. **Understand where the relevant ESI is likely to be located and take steps to actively preserve that data.**

Find more information on preservation in Chapter 3.

Understanding Information Management

Technically, information management isn’t part of the eDiscovery process. However, getting your information “house” in order has a significant impact on the quality and risk level of your eDiscovery process. Keeping your information organized (and documenting how that information is organized and archived) helps set the stage for your eDiscovery efforts.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Consider reviewing your policies and technologies in the following areas when you are looking to improve on information management:

- E-mail management and archiving
- Records management systems
- Backup strategy and retention
- File visibility, classification, and clean-up
- Retention policies

The better an organization governs its information prior to the triggering event of an eDiscovery matter, the more efficient and painless eDiscovery will be!
Chapter 2
Understanding ESI and IT Systems

In This Chapter
▶ Learning about e-mail systems and preservation
▶ Examining other potential sources of electronic data
▶ Accounting for all your electronic data

Although the next step in the process is to preserve and collect relevant information, it’s difficult because most lawyers don’t understand the basics of information technology. In this chapter, we look at a high-level overview of a few basic IT issues that are important to the eDiscovery process.

E-mail Systems
Most companies use e-mail — it’s by far the most common IT system. Beyond mere communication, e-mail can be used as:

✔ A workflow mechanism to route approvals
✔ A filing system and de facto records management system
✔ An electronic replacement for approvals of contracts, understandings, and requisitions
✔ A host of other things for which it is not really built

For these reasons, e-mail is a treasure trove of information and is one type of data that is almost always requested.

The basics of an e-mail system are actually fairly simple and can be compared to the U.S. Postal system:
An e-mail user creates a message that is addressed to another user (much like writing a letter).

The user sends this message through their company e-mail server (much like putting outgoing mail into a mailbox).

The e-mail traverses a network until it reaches its destination (much like a letter goes through the postal system until it reaches the destination mailbox).

Once in the destination mailbox, the message remains there until the intended recipient opens it (much like a real letter).

Of course, important differences exist. Among others — e-mail travels almost instantaneously, and you can send an e-mail to 200 people as easily as you can to one person. Multiple copies of e-mails can be created all along the way to the recipient, and additional copies can be kept even after receipt.

Most users interact with the corporate e-mail system through a client, an e-mail program (such as Microsoft Outlook or Lotus Notes) that provides access to messages within the larger corporate e-mail system. Many of these e-mail systems allow users to copy or move e-mails from the main corporate server (where space may be limited) to their local hard drives. Once there, end-users can maintain messages in local e-mail containers (typically PST or NSF files) that are completely under their control. These local containers can be copied, moved, backed up, written to CD/DVD, or copied onto USB drives. This makes them very difficult to find for eDiscovery purposes.

Laptops, Desktops, and Fileshares

Almost everyone uses a laptop or desktop computer in their daily work. With applications such as Microsoft Office, employees can create enormous volumes of data. In addition, computers frequently create some of their own data, storing cookies from Internet sites that have been visited, generating journal logs of activities by the user and by the system, and so on. All this information is preserved and maintained on a hard drive that stores data until it is deleted by the end-user.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Many workers also have access to a fileshare, which they may know as a home share, a department directory, or a logical name like the X drive. These repositories appear as local drives on a computer, but actually reside on the company network and are used by groups, departments, or even the entire company. The IT department generally manages these shares. Even if there is just one physical system, it may appear as many different systems to end-users.

**Document Management and Records Management Systems**

A document management or records management system puts some formality into the process of creating, retaining, sharing, and destroying electronic documents. Most lawyers know these systems as *collaboration tools*. In these systems, users must generally use a check-in and check-out process to track who created and made changes to documents, as well as how long the document is supposed to be retained. Better systems even enable litigation holds and prevent anyone from deleting the document (accidentally or intentionally) until the hold has been released.

**SharePoint**

SharePoint is a Microsoft family of products designed to enable collaboration, file sharing, and Web publishing. Its ease of use has taken end-users by storm. However, the difficulty in tracking and managing SharePoint data has created headaches for IT (and soon for legal). SharePoint sites, or collections of data, frequently spring up to support a project, and users tend to be free to collaborate and share information. After the project has been completed, the site is often left behind — it’s no longer in use, but still contains important information that can be difficult to locate for eDiscovery.

**Database Management Systems**

Databases are used to maintain vast amounts of data shared among many people and in critical systems, such as for financial accounting, human resources, and manufacturing data.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
These systems maintain data in a very formal, structured manner that makes storage, searching, and sharing the data extremely efficient. It also makes the data inaccessible without the proper tools.

You can’t just “print out” or “give” someone your database. There are enormous amounts of information in a database that the other side has no right to view, and specialized, expensive software is usually necessary to operate and access the information. Yet it can be difficult to extract the specific data that is relevant to a case, and someone may have to generate a new report or query to extract that information in a usable format.

**Social Media and Cloud Computing**

It’s impossible to predict new technologies that will be invented in the next several years, but if businesses use them, they will be important to lawyers. We close this chapter with a quick look at two rapidly developing technologies that hold great promise: social media and cloud computing.

*Social media* is data that is shared among groups of people, mostly for social purposes but increasingly for both social and business uses. Current examples of social media include services like Facebook, Twitter, and LinkedIn. Because social media usually resides outside of the company’s control, it can be difficult to collect and hold — if a company is even able to identify that it is relevant to a case. In addition, its mixed business and personal usage makes corporate policies and relevance difficult to determine in many cases.

*Cloud computing* refers to delivering computing and data storage in a “power plant” model. In a cloud model, a third party is responsible for maintaining and provisioning computing services on an as-needed basis to end-users. The cloud can create compliance and eDiscovery issues because although the data is created and owned by the company, a third party is now contractually responsible for operating the infrastructure and thus may have actual control over the data.
Chapter 3
Finding, Preserving, and Collecting ESI

In This Chapter
▶ Creating your data map
▶ Getting everybody involved in the preservation and collection process
▶ Choosing the right preservation and collection methods

Now that you know a little bit more about the systems where your ESI may be stored, we cover how to actually find, preserve, and collect that data.

Finding Your Data

Before starting on any journey, it is always nice to have a map to guide you where to go. Many people embrace this concept in eDiscovery by creating a high-level data map of ESI repositories. Creating a useful data map requires perspective from both IT and legal.

A good data map isn’t the same as an IT infrastructure diagram. Instead, a data map is more of a checklist that contains high-level information about the repositories of data in your environment that are likely to have data subject to eDiscovery requirements. Look for systems outside your company that may have important information, such as customer and sales tracking systems and human resource and payroll information.
When creating a data map, begin with ESI repositories containing content that has been needed in your recent cases and add the systems that your key players access. You can even flip this idea around — ask which systems, applications, and data stores don’t contain responsive data based on your typical legal claims and regulatory requests.

Get the Right People Involved

Many companies performing eDiscovery today still maintain that identifying, preserving, and collecting data is a job that belongs to IT. That thinking is outdated for the following reasons:

- More courts mandate that legal counsel (both in-house and law firm counsel) take an active and informed role in the eDiscovery process. Failing to do so can lead to spoliation claims and sanctions against both the company and counsel individually.
- IT can’t effectively preserve and collect ESI without important input from legal on what needs to be found. Without legal input, the result is typically an over-collection, because IT preserves everything possible.
- If this step goes wrong and data isn’t preserved, the rest of the process can’t fix the initial problems.

Choosing the Right Method of Preservation and Collection

Based on your data mapping and your early case assessment, you should have a good understanding of the important data for your case. At this point, you must decide how to preserve and collect the data. These choices vary depending on your assessment of risk, the availability of resources and tools to assist you, and the timeframe for completing the work.

There are many ways to preserve and collect ESI. Table 3-1 provides a brief description of some of the most popular methods.
<table>
<thead>
<tr>
<th>Method</th>
<th>Pros</th>
<th>Cons</th>
<th>Bottom Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backup tapes</td>
<td>Easy and simple; low risk of spoliation</td>
<td>Overcollection of data resulting in huge costs; risk can actually be increased due to potential loss of restoration; can't analyze data</td>
<td>Use only as a last resort</td>
</tr>
<tr>
<td>Forensic images</td>
<td>Complete; preserves authenticity/admissibility; retains deleted/slack space</td>
<td>Tricky and expensive to do right; significant over-collection of data; no access to data</td>
<td>Use when needed: criminal cases, FCPA, allegations of deletion</td>
</tr>
<tr>
<td>Archives</td>
<td>Great for information management purposes</td>
<td>Takes time to be of value; no value outside that repository</td>
<td>Great for one specific repository</td>
</tr>
<tr>
<td>Custodian-based hold</td>
<td>Easy to do, inexpensive, can appear valid</td>
<td>Can't be relied on; heavily criticized in cases</td>
<td>Risky, use only in very limited circumstances (or as an additional layer)</td>
</tr>
<tr>
<td>Custodian-based collection</td>
<td>Easy, simple, can appear valid</td>
<td>Easy to challenge, metadata changes, can be incomplete</td>
<td>Use only for limited sources (for instance, USB drives, paper)</td>
</tr>
<tr>
<td>In-place indexing and/or collection</td>
<td>Fast, efficient, complete</td>
<td>Initial purchase required; ongoing maintenance and expertise; infrastructure overhead</td>
<td>Best practice for repetitive litigants</td>
</tr>
</tbody>
</table>

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Most litigants preserve data by holding it in place or making a copy and storing it someplace safe. Holding data in place can be fast and efficient and save money on storage. However, most systems don’t enable a real “hold-in-place” capability, or they use a system that can be easily worked around by end-users. Affecting a hold by making a copy of data works well for those who believe that collection will eventually need to be done. In addition, this method ensures that data will be preserved and maintained under a watchful eye.

Using cheap sources to preserve ESI (for example, backup tapes) can cost a lot more over the entire life of the case and substantially increase your overall risk.

**Collect What You Need**

Lawyers tend to be risk-averse, and in eDiscovery, they tend to want to collect everything. This may have worked back in the good old days of paper, but when one gigabyte is the equivalent of 77,000 pages, it just isn’t a financially feasible process.

Get your outside counsel on board! Make them read this book. Get them comfortable with your new process and make them as eDiscovery savvy as you.

CDs, server backup tapes, e-mail archives, and other data sources can contain a massive amount of information that needs to be examined, retained, or tossed out. Take a look at Table 3-2 for some common sizes.

### Table 3-2 Sizes of ESI Repositories

<table>
<thead>
<tr>
<th>ESI Repository</th>
<th>Common Size</th>
<th>Physical Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>650 MB</td>
<td>50,000 pages</td>
</tr>
<tr>
<td>E-mail PST</td>
<td>400 MB</td>
<td>3,500 e-mails (1.5 pages per e-mail)</td>
</tr>
<tr>
<td>USB drive</td>
<td>1 GB</td>
<td>77,000 pages</td>
</tr>
<tr>
<td>Laptop drive</td>
<td>40 GB</td>
<td>3 million pages</td>
</tr>
<tr>
<td>Backup tape (Super DLT)</td>
<td>60 / 120 GB</td>
<td>4 to 9 million pages</td>
</tr>
</tbody>
</table>

Chapter 4
Reviewing, Analyzing and Producing ESI

In This Chapter
▶ Reviewing all electronic information
▶ Reducing your data and your costs
▶ Understanding available production formats

After ESI has been collected, the next step is to analyze the data in the context of how it affects the case. You might also mark or tag documents that may be privileged, confidential, non-responsive, or responsive to particular issues in a case or even specific Requests for Production. In this chapter we show you how that work is currently being done, and some best practices for improvement.

Reviewing the Review Process

In the ancient days of litigation, when lawyers dealt almost exclusively with actual paper documents, each document was reviewed by someone and tagged, boxed, or stacked according to that process.

With the cost of legal review at about $18,000 to $35,000 per GB of ESI, it’s clear that handling the review in the same way as paper is incredibly expensive!
Reducing Your Data Sets

Reducing the amount of data that requires review can save a lot of money. There are many ways to cut back on data — starting with not collecting unnecessary data. Another common method is to eliminate duplicate files and e-mails.

A great tool for identifying duplicate information is known as a hash algorithm. A hash algorithm takes an electronic file as input and returns a unique sequence of data as an output. Identical files will always return the same value — but if even a single bit is different, the algorithm will return a completely different value.

You can leverage this information to ensure that duplicate documents are tagged the same way and that the document is reviewed only one time. Because it is a unique identifier, the hash value can even replace the old-fashioned and outdated Bates stamp!

Privilege review – an ounce of prevention

Another area undergoing rapid change is in reviewing information for attorney-client or work product privilege, which typically is not produced. Mistakenly producing privileged documents can result in a waiver, so parties frequently spend a lot of money and effort reviewing all their documents to determine which are privileged. In fact, having a diligent process is a significant factor in protecting against a waiver.

To protect privilege, use special agreements, or clawbacks, with the other side and review the protections under Fed. R. Evid. 502.

New technologies and processes can cut costs and save time in privilege review, including:

- **Objective privilege**: Documents are marked as initially privileged based on an objective basis (such as date range or author/sender), a privilege log is generated, and the presumption can be challenged.

- **Machine review**: An initial batch of privileged documents is marked and then similar documents are identified through computational methods.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Leverage technology to tag and manage privileged information, then produce an appropriate privilege log of those documents. Even if mistakes are made, the combination of diligence and clawback agreements can help to protect the privilege.

**Analysis – Remembering Why We Are Doing This Work**

In the midst of an eDiscovery process, there is frequently so much concern about locating, preserving, and collecting data (and preventing spoliation) that people forget the work is being undertaken for a specific purpose — to help to prove or defend against allegations made by another party. Analysis represents the stage at which you see how the data impacts a case.

Ensure that your eDiscovery process will permit you to begin analysis at the earliest possible time!

**Producing Your Information**

Production is something that you should think about from the very beginning of the case. Take a quick look at some of the most common formats in Table 4-1.

<table>
<thead>
<tr>
<th>Table 4-1 Common Production Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Production Type</strong></td>
</tr>
<tr>
<td>Native</td>
</tr>
<tr>
<td>Quasi-native</td>
</tr>
<tr>
<td>Image</td>
</tr>
<tr>
<td>Load files</td>
</tr>
<tr>
<td>Hosted</td>
</tr>
<tr>
<td>Paper</td>
</tr>
</tbody>
</table>

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
In general, don’t use a single form of production for all items. Instead, tailor the form to the data. With well-established tools requiring TIFF and/or load files, those formats may be needed, and they’re good for content with simple text like e-mails or Microsoft Word documents. Native forms are generally best for spreadsheets, CAD, and other complex data sources. However, these sources can’t be Bates stamped in the traditional manner. Quasi-native forms can also work well for sources such as e-mail and databases.

Understand current downstream processes and vendors and leverage tools behind the firewall to feed downstream processing in a format that suits the purpose.

Native Format versus Images

A native file usually means a file that is saved in a format that only certain programs can recognize. For example, a Microsoft Word document is automatically saved with a .doc extension. Data saved in a program’s native file format will store everything that you have created with the program.

When it comes to leveraging technology in eDiscovery, a major mistake is to convert native ESI (with all its metadata, text, and other information that you can leverage) into an image format such as PDF or TIFF, which is basically just a picture of how the document looks. Frequently, this conversion is done because lawyers are more comfortable with images that can eventually be printed out onto paper. Other times, images are used solely so that the electronic equivalent of a Bates stamp number can be affixed to the document. Maintaining data natively, for as long as possible, is the best way to preserve metadata, search for information, and keep your production options open!
Chapter 5

Ten Ways to Prepare Your Legal Department for eDiscovery

In This Chapter
▶ Getting to know everyone
▶ Being flexible

Think of this section as a tip sheet you can use to refresh your memory. These are important points to keep in mind, so check early and check often.

✔ **Embrace the technology:** eDiscovery means you have to be familiar with all kinds of technology. You don’t have to be an expert, but be conversant with the technology and know the terms and concepts. We promise it won’t hurt you.

✔ **Introduce yourself early:** Way before eDiscovery becomes a necessity, go down and meet the IT team. Let them know how you want to work together and what kind of goals you want to achieve. A little proactive work goes a long way.

✔ **Filter it down:** Dumping a ton of information on any eDiscovery process bogs down the process and frustrates everybody involved. Diligent de-duplication and filtering makes sure you only get the information you (and others) might need.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
 vier Know the process: If you make yourself familiar with the EDRM and follow it diligently, you'll save yourself a ton of hassle and possible sanctions. Time to do your homework.

✓ Be flexible: Not all discovery processes follow the same steps and involve the same data sources. Be open to locating all possible information and make sure it gets included in the process.

✓ Remember the data about the data: Where and when an e-mail originated can be just as important as what’s in the e-mail — as is where the e-mail was accessed and how it got to that device. Be sure you consider all the available metadata.

✓ Look outside the company: Data tends to get out from behind firewalls and company-approved servers and devices. Be sure to check any applicable social media accounts, cloud computing accounts, and other outlets for information about your litigation.

✓ Read the map: Work with your IT department to develop a clear, reliable data map. Knowing where all the data resides makes any eDiscovery process much easier to manage.

✓ Watch out for privileged information: When turning over data, be sure to know where privileged information originates and make sure it doesn't get included in data turned over to other parties. A large volume of information makes this effort difficult, but it's still a necessary process.

✓ Use multiple formats: Evidence goes beyond paper. Anticipate the most useful information formats for the case and use multiple formats if necessary to share all applicable information.
Learn to:

• Work well with the legal team
• Understand common legal terms and processes
• Understand your impact on the process

Jim Shook, Esq.
Heidi Maher, Esq.

The Technologist’s Guide to eDiscovery Law
FOR DUMMIES

Compliments of EMC²

The Lawyer’s Guide to eDiscovery Technology

Compliments of EMC²

Dummies
Addressing Today’s eDiscovery and GRC Challenges

Governance, Risk, and Compliance regulations surrounding discovery of electronically stored information have changed corporate business practices. With expanding rules and regulations, the legal department must defend the corporate interest, meet governance challenges, and ensure regulatory compliance, while balancing risks and expense. To meet today’s challenges, corporate attorneys need legal knowledge and experience, as well as an increased understanding of corporate IT capabilities, policies, and technologies. At the same time, corporate IT needs to understand the significant risk and cost implications of implementing technologies that do not address legal, compliance, or regulatory requirements.

The authors of this book are a part of the EMC Information Governance, Risk, and Compliance Practice, which is a specialty practice within EMC Corporation, consisting of seasoned attorneys who also have backgrounds in business law, eDiscovery, information technology, records management, security, and compliance. The team helps organizations understand their legal and regulatory obligations, best practices to meet those obligations, and how EMC solutions can be leveraged as a part of a GRC or information governance strategy and process.

For more information about the EMC capabilities and solutions, visit www.emc.com/informationgovernance or contact us at goodgovernance@emc.com.
The Technologist’s Guide to eDiscovery Law
FOR DUMMIES


These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Introduction

The world continues to create and store data at incredible rates. In 2009, the IDC corporation says the total amount of the world’s digital information grew to 800,000 petabytes, equivalent to a stack of DVDs reaching from the earth to the moon and back. The number is expected to grow to 1.2 million petabytes, or 1.2 zettabytes by the end of 2010, and then to 1.8 million zettabytes by 2011.

A petabyte is a thousand terabytes or a million gigabytes.

The bad news for companies like yours is that IDC also says that enterprises are liable for about 80 percent of that data — and that data falls into your custody. It’s an enormous task further complicated by the fact that an amazing web of compliance and privacy laws and requirements demand that chunks of that data be retained, deleted, stored, or secured, often in a conflicting manner.

And you’re not a lawyer — you work in IT. Your job is to corral and store all of that data, right? True, but it also involves producing that data whenever required, and that includes legal proceedings. You have to furnish everything, from the actual information to logs and metadata regarding that information.

In the civil litigation system in the United States, companies must frequently locate and preserve data that is relevant to a specific case. It can be a daunting task given the volume of information and number of repositories that may hold relevant information. Companies unprepared for this process frequently face a difficult choice: Spend a lot of money to fully and properly respond to this challenge; or just wing it and hope for the best.

If things don’t go well, your company can face serious sanctions from a court and comparable losses in the stock market and the court of public opinion. You may not have to deal with that directly, but your executives will, and that means you’ll deal with it by proxy.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Because of the high cost and risk, many companies have undertaken initiatives to improve their eDiscovery processes. Even then, it can be difficult work because it requires smooth communication and coordination between two groups that frequently just do not seem to speak the same language: legal and IT.

In fact, without personnel who can understand both legal and IT issues, initiatives frequently falter or fail.

**About This Book**

This book is written to help personnel in both the legal and IT departments to understand those parts of the eDiscovery process that may not come naturally to them. Hopefully, what you read here will help to bring the two sides together in blissful eDiscovery harmony.

The side you’re reading right now assists IT professionals understand the legal pitfalls and requirements of the eDiscovery process. The other side of the book helps lawyers try to grasp the technology involved in the eDiscovery process. They may still call you with questions just because that’s what they do, but we try to explain stuff the best we can. On the off-chance that these two groups are fighting over a copy of this book, please either tear it in half or call EMC for another complimentary copy.

**Icons Used in This Book**

This book uses a couple of icons to draw your attention to special information. You’ll gain some good knowledge if you stop and take notice.

- **Tip**
  - This icon denotes information that will save you time or money.

- **Warning**
  - This icon points out common traps and pitfalls you might run into. Pay attention to these bad boys — it’s better to be warned now that run into trouble later.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Chapter 1
Understanding the Litigation Process

In This Chapter
► Understanding what kinds of litigation require electronically stored information
► Learning about the rules and regulations governing this process
► Discovering your role in the process

For many raised on lawyer television shows (from Petrocelli to LA Law and more recently, The Practice), it comes as a shock that there’s a lot more to the civil litigation process than heading immediately to a trial. In fact, only a very small percentage of cases (around 1 percent in Federal Court) actually wind up in trial. Cases frequently settle, or they’re won or lost on motions well before a trial begins. Knowing the basics and having some understanding of the full litigation process is a significant asset in working in eDiscovery.

A Civil Lawsuit
A lawsuit is a complex process that we break down into a few basic components.

The Complaint: Starting the lawsuit
The Complaint is the first document in the litigation process, and it contains the basic allegations upon which the plaintiff (the party starting the process) is making a claim against the
defendant (the party against whom the suit is filed). The complaint is actually one of several pleadings (formal statements filed with the court) that ensure everyone is aware of the claims being made.

Responding to the lawsuit

The defendant usually has less than a month to file an Answer (another pleading), which is a formal document that responds to the allegations in the complaint. In general, the Answer responds to the claims and factual statements made by the plaintiff and sets up any defenses. In some cases, the answer may contain counterclaims, or claims against the plaintiff. For example, the defendant sued for breach of contract might say that it is the plaintiff who actually breached the contract.

Crossclaims are filed when parties have claims against each other, such as when one of several defendants claims that another defendant is obligated to reimburse it for any damages. New parties can even be brought into a case if needed to administer justice.

Discovery: Getting to the truth

During discovery, the parties must provide each other with information in their possession related to the claims in the lawsuit. They may request other information from each other, and they may even ask third parties for information that is relevant. They have many tools to do this, including the following:

- **Depositions:** Lawyers have witnesses respond orally to questioning, under oath, with everything spoken being recorded by a court reporter.

- **Interrogatories:** Lawyers ask written questions that must be responded to, in writing, and under oath.

- **Requests for production of documents and things:** Lawyers file written requests from one party to the other asking for documents (including electronic information) and other physical objects that may be relevant to a case. For example, the defendant in a contract case might ask for all copies, including draft copies, of the contract at issue.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Although *electronically stored information* (ESI) was never excluded from the discovery process, it was not until December 2006 (when the Federal Rules of Civil Procedure were amended), that actual rules were clearly established for incorporating ESI as part of the discovery process.

**Understanding the 30(b)(6) Deposition**

The 30(b)(6) deposition is a special kind of request. Most depositions require a party to specify the person that it wishes to depose. A 30(b)(6) deposition enables a party to instead specify certain topics that it wants to discuss. The responding party must then designate and supply the people within its organization that can provide that information.

Many have used this tool effectively in cases involving eDiscovery issues, where one party might compel the other to provide detailed information about its IT systems. This information can include backup and retention policies, data infrastructure, and e-mail accounts. These depositions can be very damaging if an IT deponent is unprepared or provides information about systems that don’t work as legal had assumed.

**Subpoenas**

Parties that are not named in a lawsuit may also have information relevant to the case. For example, a party may want a copy of bank information or medical records important to their case. The rules of discovery provide a mechanism to obtain this information directly from that party through a document called a *subpoena*.

Subpoenas have an important role in eDiscovery because information can be obtained directly from a source that may have information the other party hadn’t considered, such as text messages. Today, this can even include social networks like Facebook and Twitter.

If you’re responding to a subpoena, you may be able to recover costs incurred in responding. So make sure that you track your costs and give this information to the lawyers on the team.
Motion practice: Narrowing the issues

The parties also have the opportunity throughout most of the litigation process to ask the court to decide certain issues of law related to the case. If granted, some of these motions can end the case. In other situations, the motions narrow or limit the issues (or the potential damages) so that a settlement can be more easily achieved.

Motion practice is also the point when a party might claim that the other party hasn’t met its eDiscovery obligations in some way. Generally, a motion alleging spoliation (intentionally or unintentionally withholding or damaging evidence) by the other party will also request some type of sanction, ranging from a monetary fine all the way to a terminating sanction, which ends the case in the other party’s favor.

Trial: Letting someone else decide the outcome

If motions don’t resolve a case and a settlement can’t be agreed on, the matter will be tried to a judge or jury. Getting all the way to a trial is one of the most exciting aspects of litigation. However, the time and expense of reaching trial makes it extremely impractical for many litigants. In addition, a trial presents certain risks that many companies aren’t willing to take. For example, while a party may believe that a case is only worth a maximum of $500,000 under any circumstances, a jury could view the case very differently. The result could be an award of nothing or $10,000,000 — or any amount between. The uncertainty of trial causes many companies to avoid trying cases except as a last resort.

Appeal: When you don’t like the outcome

After a final verdict, a party may believe that mistakes were made by the judge, jury, or both during the litigation process and file a formal appeal listing its concerns. The other side will respond, and typically the parties will argue the appeal
before a court that is higher than the trial court. There are even a few levels of appeals (all the way to the U.S. Supreme Court) if a party doesn’t like the answer it receives on appeal. The appellate process doesn’t typically permit the introduction of any new facts or documents; instead, the higher courts review the legal decisions made in the case.

**A long haul**

The entire litigation process can take years; large cases may not be resolved for five years or more. As you think about your eDiscovery work, remember that the other side may have the right to ask you exactly what work you did — three or four years later!

**The Rules**

People working with the eDiscovery process often hear that something is required by “The Rules.” But what does that mean? Where do these rules come from? There are actually many different types of rules within each court system, and the rules can even vary among different courts.

Make sure you have been given information on the rules that will apply to your case. Even courts within the same state can have different requirements.

**Federal Courts**

Only certain types of cases can be tried in Federal District Courts. They must involve federal concerns or include parties who are citizens from different states. The U.S. Court of Appeals (Districts 1 through 11 plus the DC and Federal Circuit) will hear appeals from these courts. The U.S. Supreme Court, which only takes cases that it wishes to hear, is the final appellate court.

The Federal Rules of Civil Procedure (or FRCP) govern the process for civil cases that are tried in the Federal Courts. These rules typically govern the administrative process and are normally what are referred to in the eDiscovery process. There are many other rules, such as the Federal Rules of Evidence, that may also apply.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
State courts

Most civil cases are litigated in state courts. Decisions rendered at the trial court level are generally appealable to a court of appeals, and then (potentially) to a state supreme court. Some decisions are even appealable from there to the U.S. Supreme Court.

Each state has its own rules of civil procedure, which are usually modeled after the FRCP but can vary in important ways.

Caselaw

Whenever an important new ruling is made that affects the eDiscovery process, many people will hear about it, write about it, and talk about it. But what does such a caselaw decision really mean?

Typically, cases on eDiscovery issues are handed down by the various courts of appeals, and technically they’re effective only in the jurisdiction in which they were decided. These cases then form precedent that should be followed in that jurisdiction unless overturned by a higher court or in a later case. Every jurisdiction is free to adopt, reject, or ignore decisions from parallel jurisdictions (such as other states or even Federal Courts in other states), but they must follow decisions from higher courts in their jurisdiction.

For example, the early eDiscovery case of Zubulake v. UBS Warburg, LLC is frequently cited as jump-starting the awareness of eDiscovery issues. Technically, Zubulake is controlling only in the Second Circuit, where the Court that decided the case is based. However, because it was well-reasoned and many states had similar rules and/or laws to those that Zubulake depended on, the case was largely accepted or followed in many other jurisdictions.
Chapter 2

Starting an eDiscovery Process

In This Chapter

▶ Understanding how to map data sources
▶ Anticipating litigation involving ESI
▶ Learning what triggers the eDiscovery process

Good eDiscovery requires a blend of both legal and IT skills and knowledge. However, the IT knowledge and information required tends to be focused on different issues than what an IT professional uses in regular day-to-day activities. In this chapter, we look at how IT can support legal in better understanding systems that contain ESI. We also explore how you should identify the appropriate starting point for the eDiscovery process.

The Electronic Discovery Reference Model (EDRM) is an excellent tool for analyzing and working on eDiscovery requirements. Both legal and IT teams must understand the EDRM model, but there’s only so much space we have to discuss these issues. Refer to Chapter 1 on the flip side of this book to familiarize yourself with EDRM.

Data Mapping

A great starting point for eDiscovery (and to get legal and IT working together) is to create a data map. Before you pull out your network diagram, understand that a data map is just a useful listing of data (ESI) that is created, used, or stored and that is likely to be needed in litigation. The eDiscovery team
uses the data map to determine, at the outset of a matter, which repositories of ESI are likely to be important in a case and figure out how to hold and collect data from those repositories.

A data map has no fixed form — it just needs to be useful to the people that are using it. Table 2-1 illustrates a portion of a sample data map to provide a general idea.

<table>
<thead>
<tr>
<th>Table 2-1 A Sample Data Map</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Repository</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Laptops</td>
</tr>
<tr>
<td>E-mail server</td>
</tr>
<tr>
<td>Salesforce</td>
</tr>
</tbody>
</table>

Update your data map regularly to make sure it remains useful.

The map doesn’t need to be (and probably shouldn’t be) too detailed from an IT perspective. For example, it would be a rare case where IP addresses, network segments, and even application versions would be necessary. Keeping that kind of detail makes it less likely that the map will be kept updated!
Chapter 2: Starting an eDiscovery Process

Litigation Holds Begin before the Lawsuit!

Although smart companies prepare for eDiscovery proactively, an actual eDiscovery process is never done in a vacuum. It is always done in response to a triggering event. However, that trigger is frequently misunderstood.

The actual trigger that begins an eDiscovery process is usually not the filing of a lawsuit, but frequently something that is known about much earlier. When a party can reasonably anticipate that litigation will commence, the process officially begins and a party must begin to *hold*, or preserve, ESI that is relevant to that matter.

The legal department should be made aware of potential litigation at the earliest opportunity, so that it can make a decision on whether it’s time to preserve data.

Basically, *reasonable anticipation* means that an event has taken place that leads to a reasonable conclusion that a lawsuit will commence. When that event occurs, lawyers have the immediate duty to preserve ESI that is reasonably accessible, if it is potentially relevant. If they don’t preserve, then they and/or their company could be subject to sanctions. That’s why the legal team frequently asks IT to drop everything in order to immediately work on preserving and/or collecting that information.

Here are a few scenarios where the *reasonably likely* requirement might be triggered:

- An employee files a harassment complaint with the HR department that merits an investigation.
- A company decides to pursue litigation against one of its competitors for violating its patents.
- A malfunctioning product has caused a serious accident.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Once the eDiscovery process has been triggered, certain information and storage media (ESI, physical documents, and other objects) must be preserved so that they can be reviewed during the litigation. In today’s digital explosion, it’s fortunate that not all data belonging to a company must be preserved. In this chapter, we look more specifically at what must be preserved and the best way to save relevant ESI.

Determining What Is within Scope

Unfortunately, there’s no literal, clear-cut definition of what ESI must be preserved for a case. Instead, you must rely on a flexible rule that’s roughly equivalent to: Preserve anything that might be used as evidence or anything that might reasonably lead to the discovery of such evidence.

Because most litigation holds should begin before or very early in a lawsuit, the process can get complicated. For example, there’s a common misconception that selecting ESI for preservation can be done by using keywords to identify data. Although keywords frequently have a place in the eDiscovery process, they are not a substitute for a proper legal hold.
process, they generally can’t be used for preservation. The issues involved in a lawsuit can be broad and complex, and narrowing those issues down to a set of keywords can be very difficult. In addition, if a lawsuit hasn’t yet been filed, there is nobody on the other side to agree upon the keywords.

Determining a proper scope requires the legal team and IT to work together to understand the systems and data custodians that likely have data relevant to a case. For example, in a contract case, you may know the employees who worked on the contract. You may also have a date range that is important to the case. That information creates a good starting point from which you can begin to determine ESI that is important to preserve.

**Proportionality**

Another potential limit to the scope of ESI is the idea of proportionality. In short, proportionality means that the effort and expense of eDiscovery should bear some reasonable relationship to the “value” of a case.

Your lawyers will need your help in making proportionality assessments. Although they should understand the “value” of a case, they may not understand how much certain eDiscovery work will cost. So before you take forensic images of 1,000 desktops or order 50 new backup tapes every week to hold data for litigation, make sure that you provide an estimate of how much the work may cost.

Proportionality is a quickly evolving concept with a lot of nuances. For example, don’t assume that you will not have to spend $500,000 in a case that you think is worth just $50,000. The other party may have a very different opinion of both the value of the case and the cost of eDiscovery.

**Preserving and Collecting ESI**

Once ESI is identified as being within scope for a case, it must be preserved. Nothing in the FRCP (or elsewhere) mandates that preservation should be done in a specific way, so the
actual method for preserving ESI will vary. Weigh the risk and expense involved in each type of preservation method so that you and your attorneys choose the right one for your case.

You can review several common collection method in Table 3-1.

<table>
<thead>
<tr>
<th>Method</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hold in place</td>
<td>Simple, quick to implement, saves on storage</td>
<td>Works well only in systems designed with holds; risk of physical loss or user intervention</td>
</tr>
<tr>
<td>Forensic imaging</td>
<td>Complete preservation of all data; generally good for authenticity and admissibility concerns</td>
<td>Expensive across even moderate numbers of custodians; substantial overcollection of ESI</td>
</tr>
<tr>
<td>Custodian or manual hold</td>
<td>Easy, simple, fast</td>
<td>Substantial risk that custodians will not pay attention, not understand, or will intentionally miss data</td>
</tr>
<tr>
<td>Automated collection</td>
<td>Requires tools and incremental storage</td>
<td>Safe, fast, enables quick and easy early case/data assessment</td>
</tr>
</tbody>
</table>

**ESI Collection and Metadata**

At some point after being preserved, ESI must be collected or copied to another location in a legally sound manner. Collection enables further processing, such as de-duplication, normalization of data, analysis of the data (such as concepts and e-mail threading), and even the review and production phases.

The collection process should be carefully designed so that important metadata isn’t lost or changed while ESI is being collected. Although metadata will not be important in all cases, it can be difficult to know beforehand whether it will be required in a specific case. So create a process that preserves metadata.
What Does Legal Do with This Stuff?

In This Chapter

▶ Reviewing and analyzing the data with the legal team
▶ Assessing the case early in the process
▶ Choosing the right production format (or formats) for the case

The eDiscovery process can be so difficult and technical that people often lose sight of its purpose. The real reason for worrying about all this ESI is that there is a potential lawsuit or investigation, and the lawyers need to understand what happened so that they can put together their legal case. Although much of a case involves testimony from witnesses and physical objects, ESI is extremely important in understanding what happened in a case.

Legal Review and Analysis

The legal team will review and analyze the ESI to see how it supports or contradicts other information about what has transpired.

In addition, the legal team will deploy different methods to review data to determine whether it meets the following conditions:

✔ The data is relevant to the case and should be produced to the other side.
✔ The data is irrelevant.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
The data is confidential or subject to privilege and should not be produced, produced separately from other information so that it can be protected, or redacted to remove confidential information.

With the large volumes of data present in most cases, the review process will normally need to take a phased approach. Data that is clearly irrelevant, such as spam e-mails and information well outside of the relevant date range for the case, can be eliminated. Other information may require “eyes-on” review, where a person with legal training must actually review the data and determine its relevance.

**Early Case and Data Assessment**

*Early case assessment* (ECA) is the process of making determinations about a case while it is still in its early stages. This typically involves a high-level understanding of the facts of the case, the status of any legal issues, whether the organization has insurance to cover liability and/or attorney’s fees, and other important information.

An important part of ECA is the ability to have a basic understanding of the company’s ESI that is relevant to the case, sometimes known as *early data assessment* (EDA). In most cases, this requires knowing how much data is going to be collected and how expensive or burdensome that process will be, as well as the ability to review information maintained by the key players.

Some methods of preservation and collection will allow early EDA/ECA, while others may hinder it by precluding an early review of the data. For example, forensic images can take a while to create and then process, as does restoring backup tapes. Even custodian preservation provides little assistance for EDA. On the other hand, processes that enable legal investigators to have early access to view the data will significantly enhance early assessment processes.

Better ECA/EDA is important for many reasons. For instance, a better understanding of the actual position in a case will lead to a better settlement position. Also, making determinations without data can lead to taking positions in a case that are expensive or difficult to defend.
Analyzing and Producing ESI

Eventually, ESI must be produced to the other side in a case. Production of data can take many forms, and we look at a few here. Remember that you should have an agreement with the other side about the format in which you are producing the data.

You will also have the opportunity to specify the format in which you would like data from the other side. Help your lawyers to understand which format is best for the specific requirements of each case.

**Paper**

Lawyers love paper! It may be difficult to believe, but many litigants (or at least their attorneys) still prefer to deal with paper and just want to print everything. *Bates stamping* (a process where documents are stamped with a unique string of letters and numbers) is also frequently a part of this process so that documents can be more readily identified. Still, printing ESI to paper has drawbacks, including the loss of most metadata and an inability to electronically search. But many parties may still request this format.

**Images**

Some parties don’t want to deal with boxes (or truckloads) of paper, but still want image files or pictures of all documents. In these cases, the ESI is converted into image files such as TIF, GIF, or PDF formats. Note that once again, metadata is lost, and in straight image files, there is little ability to electronically search through text.

**Native**

A default format is to produce the data in its native format. This tends to work reasonably well for most ESI, such as e-mail and productivity files (word processing, spreadsheets). It is more difficult when you’re producing structured data, such as information from databases. But if you’re prepared to produce in native format and you have properly preserved metadata, you should be able to meet almost any request.

These materials are the copyright of Wiley Publishing, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Proprietary formats

A frequent compromise format is to generate information in certain proprietary formats, sometimes called load files. These files combine images with text so that metadata can be preserved and text can still be searched. Some of these formats are specific to the products or platforms on which they’re loaded, such as the Concordance and Summation platforms. This can be a good compromise format because lawyers can see the images and the metadata is still available, if needed. However, you must have a license to use one of these products (or your own solution must have the ability to access these formats).

Other standards (EDRM XML)

The promise of an open standard (a Rosetta stone for eDiscovery formats) is always an intriguing proposition. EDRM is currently continuing its work on such a standard (known as EDRM XML), which is gaining traction in the industry. Because the standard is meant to be common and open to all, it holds great promise for ending many of the current issues with varying formats.
Chapter 5

Ten Ways to Prepare Your IT Department for eDiscovery

In This Chapter
▶ Pinpointing your data
▶ Working together

This chapter gives you a quick reference to important steps in the eDiscovery process and what you need to do to be prepared for them.

✔ Know where your data lives: A clear and efficient data map makes it obvious where your important data is located and how you can efficiently retrieve it. Be sure to take into account all possible locations — your servers, your computers, your mobile devices, and even social media accounts.

✔ Learn the rules: It’s easier to play the game when you know exactly how to play it. Learn what regulations apply to your data (especially in the extremely regulated fields of health care and finance) and make sure you’re compliant with each and every one.

✔ Don’t be too literal: Remember that Legal and IT standards have almost nothing in common. Where IT is binary, Legal works in gray areas like “reasonableness.” Make sure that you don’t apply the wrong standard.
Keep a trail: Legal proceedings can carry on for many months or even years. Make sure that you have a record of what your team did in finding and collecting information, so that someone can talk about it later — even if that means four or five years later.

Work as a team: You’re not in this alone. The legal team needs to work with you to provide speedy, compliant answers to requests and subpoenas. Keep the overarching goals of your efforts in mind when making decisions.

Data isn’t the only data: No computer file exists in a vacuum (unless you accidentally swept up a USB drive, but that’s a different case). Be sure to preserve all applicable metadata for your files and make it available when necessary.

Choose the right media: You already know a ton of storage media used to share data with others, from .CSV files to flash memory drives to (shudder) paper. Be sure you know which one is appropriate for the situation and use it correctly.

Know when (and where) to hold ‘em: Data retention policies demonstrate a clear and approved method of retaining and destroying records. Create a valid policy and stick with it, so you can better furnish data and not be burdened with excessive data storage requirements.

Do it as early as possible: There’s no time like the present. The earlier you prepare your systems for eDiscovery requirements, the less trouble you’ll encounter down the line.

Know the standards: Whether they’re open or proprietary, standards are still standards. Find out which data storage standards your system uses and the best way to extract them in case the legal team requests them.