

# THE Data Preservation Data Collection CONTINUUM



Businesses nationwide identify electronic discovery as one reason their litigation budgets have grown inexorably year after year. In-house counsel attend seminars, listen to web casts and study brochures describing coveted technology solutions<sup>1</sup> designed to make their lives easier. Thought is given to conducting the Litigation Readiness Assessment, which they are told, and undoubtedly believe, will lead down the path to a more disciplined and cost effective Litigation Response Strategy. They may even be working on compiling the persuasive statistics to include in next year's budget request illustrating just how costly it is to recreate the electronic discovery wheel with each new litigation threat.<sup>2</sup> The reality for many in-house counsel, however, looks like a tight budget, limited IT and Human Resources, amended Federal Rules of Civil Procedure that seemingly are not making life easier and a complaint on their desk signaling the need to immediately preserve data and develop a collection strategy.

Nothing in the 2006 amendments to the Federal Rules of Civil Procedure changed the familiar common-law and statutory duties to preserve data. What the amendments did do is highlight the challenges faced in meeting those obligations when it comes to electronically stored information. Here we will not discuss the elusive issues of trigger and scope,<sup>1</sup> but rather the relationship between data preservation and data collection and how to avoid doing the right things the wrong way or perhaps failing to do them altogether.

BY ANN MARIE GIBBS, SHEILA MACKAY,  
AND DOUG STEWART\*

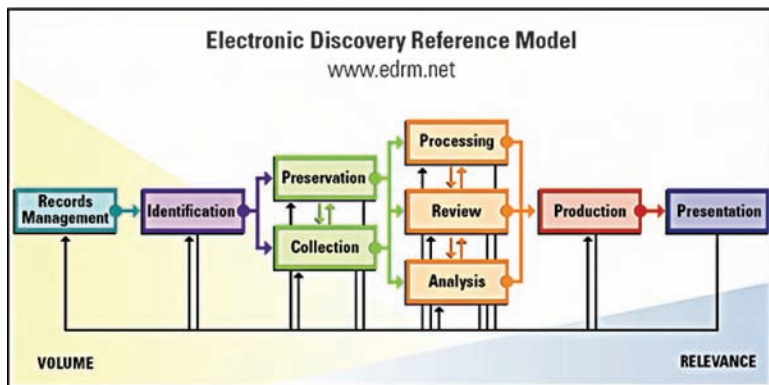
## DATA PRESERVATION: The Litigation Hold

In a typical scenario, receipt of a complaint is followed by the issuance of a litigation hold notice to employees and appropriate third parties. A monitoring protocol is put in place a la Zubulake V. (Zubulake v. UBS Warburg LLC, 229 F.R.D. 422 (S.D.N.Y. 2004).) Data custodians will be alerted to the need to retain potentially relevant electronically stored information and ensure it is not destroyed, deleted or altered. Corporate IT is hopefully reviewing routine computer operations to determine whether any settings need to be changed or policies altered to prevent the loss of data. Counsel is busy preparing initial disclosures and making decisions regarding data accessibility. A team is dispatched to interview key custodians to determine where to find the electronically stored information that is relevant to the litigation. Custodians are reminded to preserve potentially relevant electronically stored information, wherever it resides; workstation, laptop or on the thumb drive casually lying on their desk. They are provided with a contact name in the event they recall an additional data source later. Depending on the scope of the litigation, all or some of these steps were taken in a short time frame in preparation for the meet and confer during which counsel will be obliged to discuss the electronic discovery issues that impact the litigation. What next?

As seen from the model below, data preservation segues into data collection. At times these two processes run in a parallel, not linear fashion. Some data are preserved, but never collected in the sense of a physical data capture. Databases are one example of this paradigm and back-up tapes removed from rotation and stored securely, but not ultimately restored are another. Other data sources, such as key custodian hard drives, may, through forensic imaging, be simultaneously preserved and collected. The choices made at this stage regarding data collection are critical<sup>2</sup> for purposes of protecting yourself against allegations of spoliation and establishing the foundation you will

need to get your electronically stored information introduced as evidence at trial.

this analysis may prompt a party to contact opposing counsel about their pres-



## THE DATA PRESERVATION Data Collection Nexus

It is at this point in the electronic discovery process that litigants often feel like a deer paralyzed by oncoming headlights. They act diligently and in good faith to identify and protect from destruction or alteration all of the potentially relevant data that can be found at this early stage of the litigation. They identify the key custodians, those individuals who have important insight regarding the lawsuit. Now they have to gather the electronically stored information.

Before deciding on which data collection tools to use, consider ranking the different data sets identified in terms of their importance to the litigation. If you can identify key custodians, you should also be able to identify key data sets. This is a step that is often overlooked in the rush to complete the data preservation and collection phases in order to jumpstart the processing, review and analysis phases. Project outcome improves if time is taken to first assess issues related to data type, content and subject matter relevance.

The treatment afforded a critical data set known to contain the most sensitive electronically stored information of the litigation may be different than that given to less consequential data. Similarly, knowing up front that a particular data set poses serious preservation/collection technology challenges provides time to solve the problem using preferred experts. Finally, issues discovered during

ervation and collection plans prior to the meet and confer.

The more complicated the issue, the more likely a party will want to proceed with a solution that either meets with opposing counsel's approval or must be brought to the court's attention for resolution. Acting independently, with fingers crossed that no objections will be raised regarding your strategy, is a risk you should think twice about taking in a climate where courts are taking aggressive stances regarding alleged electronic discovery abuse.<sup>4</sup> From the amended Federal Rules of Civil Procedure to the current case law, counsel are admonished to meet early and often to resolve disputes. Do not make the mistake of thinking such meetings cannot take place before the meet and confer.

This is also the appropriate time to consider how the specific facts of the case and the current state of business operations will affect decisions regarding how and when to collect data. The overall data collection strategy must strike a balance between the business need to maintain productivity and the legal obligation to capture the potentially relevant data. Parties must act quickly to preserve and collect data to prevent a disgruntled employee from destroying data.

## DATA PRESERVATION: Metadata

Fundamental to an analysis of which data collection tools to use is an understanding of metadata. Although commonly referred to as "data about data",

Before deciding on which data collection tools to use, consider ranking the different data sets identified in terms of their importance to the litigation.

metadata is best defined functionally. In general it is "information associated with and made part of an electronic document that is not visible in the normal viewing or printing of that document." The two basic functional categories of metadata are:

- System metadata is automatically created by a computer system and relates to system operation and file handling (e.g., file create date and time; user name; file path)
- Application (software program) metadata can be automatically created or user generated and generally relates to application use and output generated (e.g., MS Word file properties such as track changes; modified, accessed, created ("MAC") and printed dates; Excel formulas, hidden rows, hyperlinks; email MAC and printed dates)

Before settling on a data collection strategy, first decide the extent to which you are obligated to preserve/collect metadata intact. In addition to this inquiry, also determine whether preserving original metadata is advantageous, even when preservation is not a requirement for purposes of meeting your discovery obligation. Reflect on this observation from one court:

[C]onsidering the significant costs associated with discovery of ESI, it makes little sense to go to all the bother and expense to get electronic information only to have it excluded from evidence or rejected from consideration during summary judgment because the proponent cannot lay a sufficient foundation to get it admitted. (*Lorraine v. Markel American Ins. Co.*, 241 F.R.D. 534, 538 (D.Md. 2007).)

Metadata may be part of the solution for addressing admissibility issues revolving around authenticity and the validity of a claim for a hearsay exception. Verification of the "business records" status of a

file or document will be facilitated by the capture of complete metadata. Metadata also harbor the information needed to demonstrate that data are being produced as kept in the regular course of business. Be sure to keep these issues in mind when evaluating collection options.

## DATA COLLECTION: Which Tools?

Preserving or collecting the right data the wrong way is a recipe for sanctions. Choosing the wrong tool or following an inadequate protocol to collect data undermines the considerable time and resources invested in identifying and preserving data. Select the wrong tool and you will fail to capture the data in the form you need and risk spoliation sanctions. Use the wrong protocol and an effort to verify and authenticate the data as evidence fails. Electronically stored information is dynamic; easily altered or lost. The simple act of booting up a computer or performing routine maintenance can alter or destroy the very data you are legally obligated to preserve. Different tools may be optimal for different data sets. Make sure you understand why.

## DATA COLLECTION: Methods and Tools

Understanding the basics regarding commonly deployed collection strategies is essential for purposes of choosing the strategy that satisfies legal obligations while minimizing business risks and costs. Familiarize yourself with these frequently used strategies to collect data from desktop or laptop hard drives:

## MANUAL COPY aka Drag & Drop

- Captures only active data
- Alters system metadata such as MAC dates; can lose original file path data
- Application metadata can be easily altered or contents changed if file opened for review; e.g. date last accessed, date last modified

- Custodian identifies potentially relevant electronically stored information
- Electronically stored information is copied by custodian or IT staff to a central location for collection
- Data capture cannot be verified or authenticated
- Process not auditable
- Custodians often are unaware of network locations storing file copies
- Temp files are often overlooked
- High risk of human error

## ACTIVE DATA COPY – Non Forensic Tools

- Captures only active data
- Preserves most system metadata but some file attributes may be updated; e.g. date last accessed
- Retains all application file metadata
- Includes Norton Ghost, VMWare images, Robocopy
- Generally used by IT staff to quickly replicate drive contents
- Deployed locally or remotely to capture data
- Data capture cannot be verified or authenticated
- Process not auditable
- Reports only success or failure of process
- Human risk error higher than Forensic tools

## BIT STREAM OR FULL DRIVE IMAGE – Forensic Tools

- Replicates physical hard drive bit by bit capturing active data, all metadata and unallocated space where deleted files and file fragments reside
- Includes EnCase, FTK and Logicube
- Deployed locally or remotely to capture data
- Data capture verified and authenticated with hash value
- Can be encrypted and password protected
- Auditable process with comprehensive audit trail

---

[C]onsidering the significant costs associated with discovery of ESI, it makes little sense to go to all the bother and expense to get electronic information only to have it excluded from evidence or rejected from consideration during summary judgment because the proponent cannot lay a sufficient foundation to get it admitted. (*Lorraine v. Markel American Ins. Co.*, 241 F.R.D. 534, 538 (D.Md. 2007).)

---

---

\*The contributors of this article are crucial members of the DAEGIS team. Ann Marie Gibbs, Esq. is the Director of Litigation Consulting, Sheila Mackay is the Vice President of Professional Development, and Doug Stewart is the Director of Technology. Each is an expert and active consultant in electronic discovery, forensically sound data collection and preservation.

---

- Minimal risk of human error
- Minimizes risk of spoliation

## GHOST and Swap

- Acquisition of original hard drive with active data, metadata and
- unallocated space containing deleted files and file fragments
- Performed by IT staff or third-party vendor
- Active data copy tool like Norton Ghost used to duplicate hard drive contents onto new drive
- New hard drive is installed in computer
- Preserve original hard drive

## DATA COLLECTION: Chain of Custody

Delivery of a detailed chain of custody record is also an important consideration when crafting a collection strategy. It is essential to document the logistics related to data acquisition and subsequent data transfer. Information regarding the time, place, and personnel involved in the data collection must be tracked. The chain of custody documentation also records all movements of the data as data are transferred through people to different storage locations. Accurate documentation is essential for purposes of authenticating the data. Realize also that each signatory on a Chain of Custody record is a potential witness at trial.

## DATA COLLECTION: Which Collection Team?

Given the high risks and costs associated with a compromised data collection, the decision about who will undertake this task is equally important as which methodology should be deployed. Many organizations believe it is an easy matter to task IT staff with this responsibility. At first glance, this solution is appealing based on the assumption it will save time and money while minimizing disruption and security risks. Upon reflection however, this assumption often proves inaccurate.

Unless your organization has a dedicated collections team, getting the commitment and focus required from an over-taxed, fully engaged IT staff may not be realistic. Data collection simply is not the same as data cloning. Cloning tools are used to quickly copy data to move it from one storage device to another. It does not need to be an auditable process supported by chain of custody documentation. Consequently, the average IT employee may have used Norton Ghost countless times without giving a thought to documenting the process. Such a laissez-faire approach is suitable for the fast-paced IT domain but dooms the meticulous world of a data collection specialist. Moreover, that same talented IT technician may never use forensic tools to capture data, making that person a poor choice in cases demanding bit by bit imaging.

Experienced third-party vendors, despite being an added budget line item, eliminate many of the risks inherent in using non-dedicated in house resources. They are adept at deploying a variety of tools using well documented protocols. They arrive prepared with chain of custody forms, protocols and evidence bags. A problem encountered can be readily solved using an extensive knowledge base gained from their exposure to a wide variety of challenges. If data authenticity issues arise, they serve as an objective third-party witness to testify about the processes they followed.

## CONCLUSION

If it is an unassailable data collection you want, it must be performed by a highly skilled and experienced collection specialist who follows a well documented protocol and uses a tool with comprehensive verification and audit features. For this type of collection, most organizations enlist trusted third-party vendors as a standard practice.

Recall that doing the right things the wrong way can lead to sanctions. Whatever your data collection needs may be, it is the synergy of the right tool, the right protocol and the right people, combined with good

faith and due diligence that signals to the court your obligations are met.

---

### NOTES:

1. This article does not address collection strategies for email archive users.
2. See generally 2007 Cohasset ARMA AIIM Electronic Records Management Survey White Paper ("For any organization which is the likely target of litigation or regulatory inquiries, the absence of a formal plan to respond to discovery requests must be considered an unacceptable risk, Not having such a system is a legal land mine waiting for detonation.").
3. These issues are discussed in Gibbs, Data Preservation: Timing and Scope, August, 2007 available at [www.daegis.com](http://www.daegis.com).
4. See MANUAL FOR COMPLEX LITIGATION 4th § 11.446 (Computerized data, however, raise unique issues concerning accuracy and authenticity.... The integrity of data may also be compromised in the course of discovery by improper search and retrieval techniques, data conversion, or mishandling. The proponent of computerized evidence has the burden of laying the proper foundation by establishing its accuracy.").
5. This chart was prepared by the Electronic Discovery Reference Model Project ([www.edrm.net](http://www.edrm.net)). Content is available free under the GNU Free Documentation License 1.2. Launched in May 2005, the EDRM Project was created to address the lack of standards and guidelines in the electronic discovery market—a problem identified in the 2003 and 2004 Socha-Gelbmann Electronic Discovery surveys as a major concern for vendors and consumers alike. The completed model was placed in the public domain in May 2006. Copyright 2005-2006. Socha Consulting LLC and Gelbmann & Associates. All rights reserved.
6. See *Qualcomm Inc. v. Broadcom Corp.*, No. 05-CV-1958-B(BLM) (S.D.Cal. Aug. 13, 2007) (ordering fourteen attorneys to appear to defend allegations they engaged in an organized program of litigation misconduct and concealment related to electronic discovery).