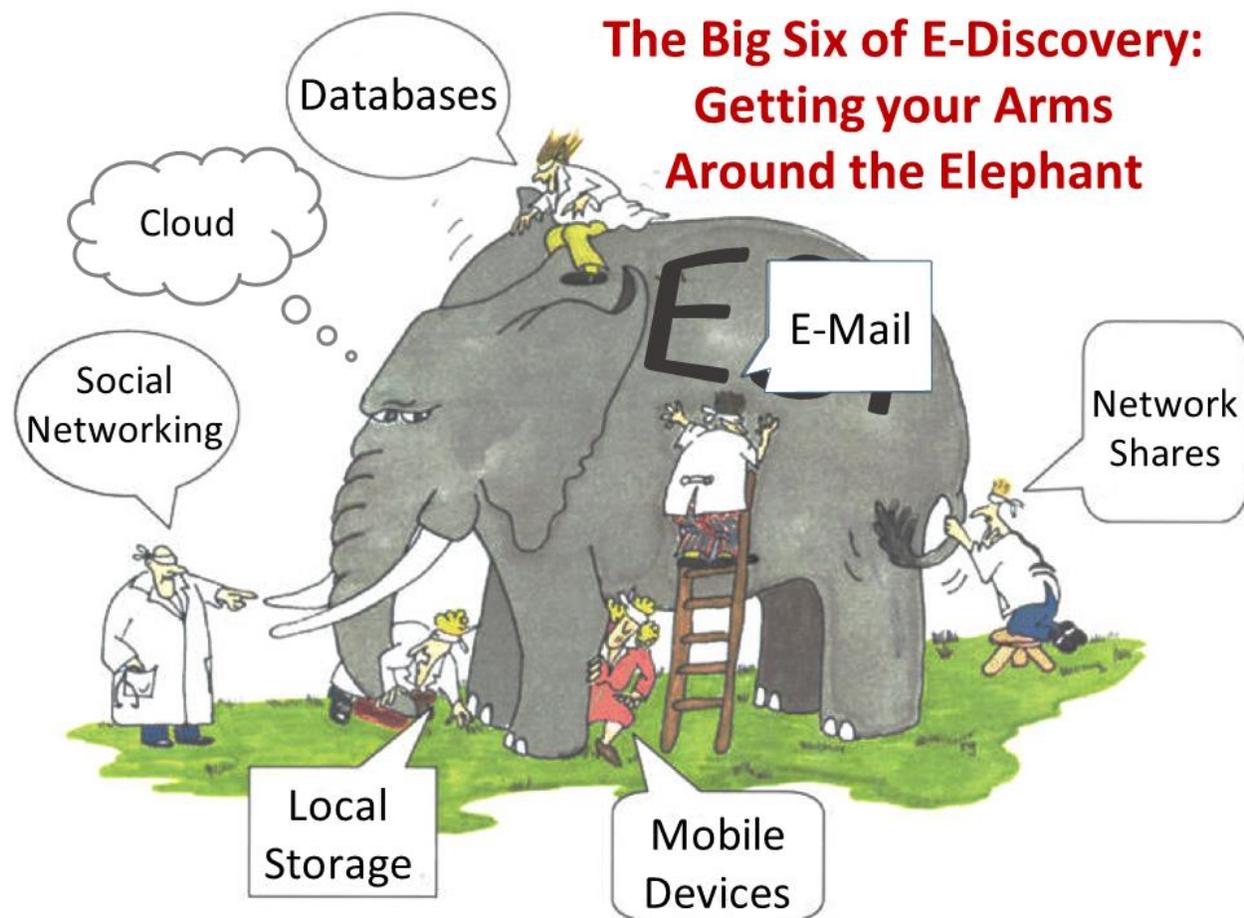


## Getting your Arms around the ESI Elephant

Many cultures and religions share the parable of the six blind men that touched an elephant. The one who grabbed the tail described the elephant as “like a snake.” The blind man who grabbed the trunk said, “no, more like a tree branch,” and the one with his arms around the elephant’s leg said, “you’re both wrong, an elephant is like a tree trunk.” The man touching the ear opined that the elephant was like a large leaf, and the blind man at the tusk said, “you’re all crazy. It is like a spear.” None of them understood the true nature of the elephant because they failed to consider all of its aspects.

In e-discovery, too, we cannot grasp the true nature of potentially responsive data until we touch many parts of the ESI elephant.



There are no forms or checklists that can take the place of understanding electronic evidence any more than a Polish phrasebook will equip you to try a case in Gdańsk. But, there are a few rules of thumb that, *applied thoughtfully*, will help you get your arms around the ESI elephant. Let’s start with the Big Six and work through some geek speak as we go.

## The Big Six...Plus

Without knowing anything about IT systems, you can safely assume there are at least six principal sources of digital evidence that may yield responsive ESI:

### 1. Key Custodians' E-Mail (Sources: server, local, archived and cloud)

Corporate computer users will have a complement of e-mail under one or more **e-mail aliases** (i.e., shorthand addresses) stored on one or more **e-mail servers**. These servers may be physical hardware managed by IT staff or **virtual machines** leased from a **cloud provider**, either running mail server software, most likely applications called **Microsoft Exchange** or **Lotus Domino**. A third potential source is a **Software as a Service (SaaS)** offering from a cloud provider, an increasingly common and important source. Webmail may be as simple as a single user's Gmail account or, like the Microsoft Office 365 product, a complete replication of an enterprise e-mail environment, sometimes supporting e-discovery preservation and search capabilities.

Users also tend to have a different, but overlapping complement of e-mail stored on desktops, laptops and handheld devices they've regularly used. On desktops and laptops, e-mail is found **locally** (on the user's hard drive) in **container files** with the file extensions **.pst** and **.ost** for Microsoft Outlook users or **.nsf** for Lotus Notes users. Finally, each user may be expected to have a substantial volume of **archived e-mail** spread across several on- and offline sources, including backup tapes, **journaling servers** and local archives on workstations and in network storage areas called **shares** (discussed below).

These locations are the "*where*" of e-mail, and it's crucial to promptly pin down "where" to ensure that your clients (or your opponents) don't overlook sources, especially any that may spontaneously disappear over time through **purges** (automatic deletion) or backup media **rotation** (reuse by overwriting).

Your goal here is to determine for each key custodian what they have in terms of:

- *Types of messages* (did they retain both Sent Items and Inbox contents? Have they retained messages as they were foldered by users?);
- *Temporal range of messages* (what are the earliest dates of e-mail messages, and are there significant gaps?); and
- *Volume* (numbers of messages and attachments versus total gigabyte volume—not the same thing).

Now, you're fleshing out the essential "*who, what, when, where and how*" of ESI.

## 2. Key Custodians' Documents and Data: Network Shares

Apart from e-mail, custodians generate most work product in the form of **productivity documents** like Microsoft Word documents, Excel spreadsheets, PowerPoint presentations and the like. These may be stored locally, *i.e.*, in a folder on the C: or D: drive of the user's computer (local storage, see below). More often, corporate custodians store work product in an area reserved to them on a network **file server** and **mapped** to a drive letter on the user's local machine. The user sees a lettered drive indistinguishable from a local drive, except that all data resides on the server, where it can be regularly backed up. This is called the user's **network share** or **file share**.

Just as users have file shares, work groups and departments often have network storage areas that are literally "shared" among multiple users depending upon the access privileges granted to them by the network administrator. These shared areas are, at once, everyone's data and no one's data because it's common for custodians to overlook **group shares** when asked to identify their data repositories. Still, these areas must be assessed and, as potentially relevant, preserved, searched and produced. Group shares may be **hosted** on company servers or "in the cloud," which is to say, in storage space of uncertain geographic location, leased from a service provider and accessed via the Internet. Enterprises employ virtual workspaces called **deal rooms** or **work rooms** where users "meet" and collaborate in cyberspace. Deal rooms have their own storage areas and other features, including message boards and communications tools--they're like Facebook for business.

## 3. Mobile Devices: Phones, Tablets, IoT

Look around you in any airport, queue, elevator and waiting room or on any street corner. Chances are many of the people you see are looking at the screen of a mobile device. According to the U.S. Center for Disease Control, more than 41% of American households have no landline phone, relying on wireless service alone. For those between the ages of 25 and 29, two-thirds are wireless-only. Per an IDC report sponsored by Facebook, four out of five people start using their smartphones within 15 minutes of waking up and, for most, it's the very first thing they do, ahead of brushing their teeth or answering nature's call.

The Apple App Store supplies over 1.5 million apps accounting for over 100 billion downloads. All of them push, pull or store some data, and many of them surely contain data relevant to litigation. More people access the internet via phones than all other devices combined. Yet, in e-discovery, litigants often turn a blind eye to the content of mobile devices, sometimes rationalizing that whatever is on the phone or tablet must be replicated somewhere else. It's no; and if you're going to make such a claim, you'd best be prepared to back it up with solid metrics (such as by comparing data residing on mobile devices against data secured from other sources routinely collected and processed in e-discovery).

The bottom line is: if you're not including the data on phones and tablets, you're surely missing relevant, unique and often highly probative information.

#### **4. Key Custodians' Documents and Data: Local Storage**

Enterprises employ network shares to insure that work product is backed up on a regular basis; but, despite a company's best efforts to shepherd custodial work product into network shares, users remain bound and determined to store data on local, physical media, including local laptop and desktop hard drives, external hard drives, thumb drives, optical disks, camera media and the like. In turn, custodians employ idiosyncratic organizational schemes or abdicate organization altogether, making their My Documents folder a huge hodgepodge of every document they've ever created or collected.

Though it's expedient to assume that no unique, potentially-responsive information resides in local storage, it's rarely a sensible or defensible assumption absent document efforts to establish that the no-local-storage policy and the local storage reality are one-and-the-same.

#### **5. Social Networking Content**

The average Facebook user visits the site 14 times daily and spends 40 minutes looking at Facebook content. That's the average; so, if you haven't visited today, some poor soul has to give Facebook 80 minutes and 28 visits. Perhaps because we believe we are sharing with "friends" or simply because nothing is private anymore, social networking content is replete with astonishingly candid photos, confessions, rants, hate speech, statements against interest and a host of other information that is evidence in the right case. Experts often blog or tweet. Spouses stray on dating and hook up sites like Tinder or Ashley Madison. Corporations receive kudos and complaints via a variety of social portals. If you aren't asking about social networking content, you're missing a lot of elephant!

#### **6. Databases (server, local and cloud)**

From Access databases on desktop machines to enterprise databases running multinational operations (think UPS or Amazon.com), databases of every stripe are embedded throughout every company. Other databases are leased or subscribed to from third-parties via the cloud (think Salesforce.com or Westlaw). Databases hold so-called **structured data**, a largely meaningless distinction when one considers that the majority of data stored within databases is unstructured, and much of what we deem unstructured data, like e-mail, is housed in databases. The key is recognizing that databases exist and must be interrogated to obtain the responsive information they hold.

The initial goal for e-discovery is to identify the databases and learn what they do, who uses them and what types and ranges of data they hold. Then, determine what standard reports they can

generate in what formats. If standard reports aren't sufficient to meet the needs in discovery, inquire into the databases **schema** (*i.e.*, its structure) and determine what **query language** the database supports to explore how data can be extracted.

## **PLUS.** Cloud Sources

The Big Six probably deserve to be termed the Big Seven by virtue of the escalating importance of the cloud as both a repository for replicated content and a burgeoning source of relevant and unique ESI in its own right. For now, it's Six Plus because it touches so many of the other six and because it's evolving so quickly that it's likely to ultimately differentiate into several distinct sources of unique, discoverable ESI. Whether we consider the shift of corporate applications and IT infrastructure to leased cloud environments like Amazon Web Services and Microsoft Azure or the tendency of individuals to store data in tools like Box, Dropbox, Google Drive, Microsoft OneDrive, Apple's iCloud and others, the cloud must be considered alone as adjunct to the other six sources when seeking to identify and preserve potentially responsive ESI.

The Big Six Plus don't cover the full range of ESI, but they encompass *most* potentially responsive data in *most* cases. A few more thoughts worth nailing to your forehead:

### **Pitfalls and Sinkholes**

Few organizations preserve all legacy data (information no longer needed in day-to-day operations); however, most retain large swaths of legacy data in backups, archives and mothballed systems. Though a party isn't obliged to electronically search or produce all of its potentially responsive legacy data when to do so would entail undue burden or cost, courts nonetheless tend to require parties resisting discovery to ascertain what they have and quantify and prove the burden and cost to search and produce it. This is an area where litigants often fail.

A second pitfall is that lawyers too willingly accept "it's gone" when a little wheedling and tenacity would reveal that the information exists and is not even particularly hard to access. It's an area where lawyers must be vigilant because litigation is regarded as a sinkhole by most everyone except the lawyers. Where ESI is concerned, custodians and system administrators assume too much, do too little or simply say whatever will make the lawyers go away.

### **Lather, Rinse and Repeat**

So long as potentially responsive data is properly preserved, it's not necessary or desirable in a high-volume ESI case to seek to secure all potentially relevant data in a single e-discovery foray. It's more effective to divide and conquer. First, collect, examine and produce the most relevant and accessible ESI from what I like to call the über-key custodians; then, use that information to guide subsequent discovery requests. Research from the NIST TREC Legal Track proves that a two-tiered e-discovery effort produces markedly better results when the parties use the information gleaned from the first tier to inform their efforts through the second.

In a bygone era of e-discovery, Thomas Edison warned, “We’ve stumbled along for a while, trying to run a new civilization in old ways, but we’ve got to start to make this world over.” A century later, lawyers stumble along, trying to deal with new evidence in old ways. We've got to start to make ourselves over.