INNOVATIVE DRIVEN

Alternate Forms of Electronically Stored Information

Alternate Forms of Electronically Stored Information

Be Aware and Beware

By Jonathan Swerdloff, Consultant and Data Systems Specialist, Innovative Driven.

Introduction

With the proliferation of new technologies, counsel needs insight into the Electronically Stored Information ("ESI") that their organizations possess and control. Not every source of ESI will be obvious at first, which is why now is the time to start investigating. Historically, lawyers dealing with their company's information faced warehouses of documents; next, there was email and unstructured data. Later, social media sites like Facebook and Twitter captured the attention of the legal profession as they presented new issues and spawned several ethics opinions1 as well as spoliation claims.2 Just as lawyers are coming to grips with social media, new sources of ESI are coming online, while the older sources are becoming even more accessible and popular. Data is everywhere, both literally and figuratively. In discovery, this creates high costs for compliance as well as high risks of inadvertent noncompliance.

In litigation, rule 26(b)(1) of the Federal Rules of Civil Procedure governs what is discoverable. The rule includes evidence related to "any non-privileged matter that is relevant to any party's claim or defense."3 This is an extremely broad set of categories, and correspondingly, often involves large volumes and varieties of data.

In preparing for litigation or investigations, it is important to know what sorts of ESI clients have been utilizing and creating. The best way to address that is to create a data map before litigation strikes. A data map is a document or series of documents that identify the who, what, and where of the ESI held by an organization. This is distinct from but related to a system map created by an IT department, which is concerned with the where and how of the organization's data infrastructure. There will be some overlap between the two, but it is important in a legal context to identify more than the mere existence of ESI: lawyers must able to make meaning of the map and have ready access to the underlying data. A good data map contains not just "traditional" sources of ESI, such as email, social media accounts and shared drives, but also many new and diverse sources of data. It should be noted that without doing diligence (such as by surveying employees), some of these data types may be unknown and invisible from an IT perspective, but could nevertheless be required for production during discovery.

Alternate Forms of Electronically Stored Information

Structured data is any stored information that follows a set of rules and stores data within identified fields. Structured data follows a data model that explains the architecture of what is stored and the relationships among fields. The data model is a separate document that explicitly identifies the elements of a structured data system, indicating what is stored in the identified fields and how they relate to one another. A data model is frequently designed to be understood by non-technical people and is used to demonstrate the business uses and interplay of systems. This is usually a visual representation like a flowchart. Specific structured data systems themselves often also have a "data dictionary" which identifies fields within a system, the relationship of those fields, and what the fields represent. Using these resources together, a general picture comes into focus as to what systems store and how systems within an organization relate to one another.

Structured data is hardly new but, until recently, it was rarely addressed in discovery contexts. Unlike unstructured ESI, structured data poses specific problems and has long been seen as difficult to access. While the data may exist, it may change in real time, the view of it may depend on the user, and the data itself may have entirely different meanings depending on how it is viewed. The rise of structured data sets that are too large for traditional data processing applications, (so-called **"Big Data"**) means that structured data will likely become increasingly important in large-scale data analysis and thus in litigation and investigations. While large and sometimes difficult to manage in discovery, structured data sets can also be the key to efficient discovery by both parties. When both sides can agree that specific evidence can be found through a particular query of structured data, it may be possible to avoid additional discovery related to that evidence.

Not all data fits into an easily identified box. In addition to structured data, there is semi-structured data, which follows a set of rules but not a set of field limits. One example of semi-structured data is **chat logs**. Chat logs are records of instant messaging applications, and follow a set of rules, identifying the username of each conversant, text of the conversation, the times of each message, and in some instances, an IP address; however, the messages themselves have attributes of unstructured data. Chat logs can be particularly important in the context of financial services, for which retention and reporting are governed by the Sarbanes-Oxley and Dodd-Frank Acts,4 and where tools such as Bloomberg chats may be the favored method of communication among employees. In some messaging tools, chat logs may or may not be created, based on user settings. For example, Google's messaging tool "gChat" has an "off the record" feature that prevents creation of a chat log.5 In other cases, chat logs may be created by third party chat clients6 that allow end users to save chats even when the "off the record" feature is turned on. A good data map can identify those users who use chat, whether logs are created, and where those chats are stored, allowing for easy collection and analysis should litigation arise.

Alternate Forms of Electronically Stored Information

Increasingly, structured data is being created by **Connected Devices**, which have sensors and the ability to trigger actions based on those sensors. The proliferation of connected devices is creating a world of data about everything, which is sometimes referred to as **The Internet of Things.** At last count, there were more than 4100 Connected Devices7 commercially available. An early question to ask in assessing where data resides from these systems is whether a particular device is "smart" and can make its own decisions, or is "dumb" and decides based on data provided from elsewhere. If it is smart, it is likely to store data, either on the device or remotely. If it is dumb, it is likely to be reacting to something else that has stored data elsewhere. This information should help you to determine where that data resides.

Wearable Devices such as the Fitbit, Apple Watch, or Google Glass are examples of connected devices, and are increasingly popular. These are devices that a user carries on her person, and may record the person's behavior, including movements, amount of activity, and even altitudes of the wearer throughout the day. These wearable devices may contain data themselves, and also frequently share that data with a central server or application, leading to potential data duplication issues.

These devices are potential sources of evidence.8 Many of them capture medical-related data, which may raise privacy concerns. Organizations should be aware of whether employees are using these devices in the workplace, and whether they might be accessing or copying data from these devices using company hardware or servers.

While connected devices and the Internet of Things can cause an influx of data and a growth in retention issues, the good news is that with a good data map, it should not be overly burdensome to deal with in discovery. Most data will not be relevant to a particular case, and it will be structured. As long as attorneys can quickly determine where relevant data resides and in what form, it can be collected and analyzed by knowledgeable experts.

The data captured by personal **mobile devices** is also on the rise. Smartphones, which contain not only cameras but sensors, are increasingly common. A photo taken on an iPhone contains more than 20 types of metadata alone, including GPS data with latitude and longitude as well as altitude. It also contains a date stamp which is usually synced to a central server's clock, the combination of which gives potential evidence even more persuasive power.

The evidentiary issues that are likely to come up with connected devices can be complex. The correct data set to look at may not be obvious without investigation. The data on a device may be required evidence, but it may be synced with a server as well. Assessing what question needs to be answered will often reveal which data is most relevant. Not every piece of data on a refrigerator is going to be important to litigation or an investigation; knowing that there were two days when the temperature was wrong and ingredients could spoil may be all that is needed. It is important to determine exactly what is at issue before identifying what to ask for or provide.

Alternate Forms of Electronically Stored Information

Voicemails and Audio Recordings pose particular issues as there is not yet a gold standard for their collection, analysis, and review. For many years voicemail was ignored or avoided by stipulation because it was deemed not reasonably accessible. More recently, VOIP phones record digital voicemail files and email them to an address associated with the call's recipient. Gone are the inaccessible magnetic tapes, replaced by yet another easily collected digital file. The difficulty is now in volume and searching, rather than collection. Automated deletion of voicemail makes this process even more complicated. There is no current technology that allows for the easy analysis of voicemail without human interaction. Voicemail and audio recordings remain cumbersome and potentially expensive, yet likely important in a discovery context.

Today's nontraditional sources are tomorrow's standards. In 2004, webmail was interesting and new while today, web-based mail is indispensable to individuals and companies alike. What does all this really mean? Litigation readiness requires a data map that reflects complex issues involved in these alternative data types. Where is the data stored? Is a device or a server "of record"? What data is captured during syncs and could any be lost? How will data be accessed? How is the relevant data within a device found? Who controls the data? Who will be an expert to explain the new form of data clearly to a court?

Automated systems have the ability to create a huge volume because they can work 24/7 creating, analyzing, and aggregating data. The implications for Information Governance and eDiscovery practice can't be escaped – the sooner an organization grapples with these issues, the better off it will be. Although the playing field has changed, the rules of the game have not. An organization must be able to identify its relevant information, whether it is a series of emails or the GPS tracker in a company vehicle. Policies and practices that include updating data maps and litigation readiness plans on a regular schedule are more important than ever. Lines of communication between Procurement, IT, and Legal must be open and forthright. Proactively identifying what an organization creates, stores, and retains provides a competitive advantage in litigation and investigations by providing documentation which support defensible decisions.

Alternate Forms of Electronically Stored Information

ABOUT INNOVATIVE DRIVEN

Innovative Driven. developer of the ONE integrated e-Discovery platform, is a leading provider of customizable e-Discovery solutions and services across the Electronic Discovery Reference Model, as well as comprehensive computer forensics and expert consulting services.

For more information contact:
Wynter Grant, Chief Revenue Officer
wynter.grant@innovativedriven.com I 877.637.4836

ABOUT THE AUTHOR

Jonathan Swerdloff, is a consultant at Innovative Driven. Prior to joining Innovative Driven, Jonathan was a litigation associate at Hughes, Hubbard & Reed LLP with over 10 years experience that included substantial eDiscovery experience that included managing large discovery projects, analysis of enterprise systems, and investigations into nontraditional data sources.

References:

1. C.F. http://www.nysba.org/Sections/Commercial_Federal_Litigation/Com_Fed_PDFs/Social_Media_Ethics_Guidelines.html

2. Painter v. Atwood 2014 WL 1089694

3. The proposed changes to this rule, expected to be effective December 2015, will add that evidence must be relevant "and proportional to the needs of the case." While it is hoped that courts will begin to give greater consideration to proportionality than they have under current Rule 26(b)(2)(C)(iii), this change will not affect parties' duties to identify all potential sources of evidence. If anything, a party may need to identify sources earlier in order to assess potential costs and relevant in context of the case.

4. Dodd-Frank created a new section of the Securities Exchange Act, Section 15f(f), which requires the retention of all written or verbal communications pertaining to trades, making this especially important in the financial sector.

5. A data map can uncover whether chat applications are being used without log creation, which could be desirable for some organizations, while for others, it could run afoul of regulatory requirements or internal document retention policies.

6. E.G. Adium

7.As of the date of this publication, there are more than 4100 devices listed at http://devices.wolfram.com/

8. In a groundbreaking first, a Canadian lawyer recently introduced Fitbit data to demonstrate that his client's ordinarily active lifestyle had become more sedentary. http://www.forbes.com/sites/parmyolson/2014/11/16/fitbit-data-court-room-personal-injury-claim/