im   ☰ Menu

Stuck in Rewind? Dynamic E-Discovery for Cloud Data

Written by Tim Anderson on January 22, 2019. Posted in Cloud, Discovery / eDiscovery.

An array of new cloud-based digital sources has emerged across the corporate landscape: chat tools, collaboration platforms, cloud productivity suites, and more. Programs like Slack, Office 365, and Salesforce bring many new and exotic challenges to corporations trying to organize, control, and produce data from these programs, and e-discovery can be particularly daunting when wrestling with the unique characteristics of this data.

In many cloud-based platforms, documents are saved approximately every 30 seconds to ensure that the user's working data is never lost. What results is a string of versions from the original save to the final copy, requiring anyone looking at the export of that data to essentially rewind the document and see each iteration across its life cycle.

Stuck in Rewind

Just as in the days of having to rewind a VHS tape to find the right point in a video, it is time-consuming and frustrating to parse through numerous versions of a cloud-based document to find the one that is relevant to the time frame of a matter. It also introduces a range of stumbling blocks across the discovery phases of collecting, processing, analyzing, and reviewing the information.

Because data is stored in cloud systems this way, documents collected from Office 365 or other cloud sources are essentially dynamic databases, making it increasingly difficult to apply standard workflows and produce accurate results in these types of scenarios. Teams working under tight deadlines to produce a dataset from these sources are increasingly challenged to determine how to pull the information from the cloud and navigate the dynamic and fluid nature of that data. Many are finding themselves stuck with no option but to slowly rewind from document to document.

Breaking Free

Fortunately, advances in e-discovery technology and processes are emerging to address some of the issues brought forth by broad cloud adoption. Teams dealing with e-discovery requests for cloud-based data will benefit from creating standard workflows that specifically navigate the nuances of cloud data sources. This includes leveraging application programming interfaces (APIs) and implementing cloud data connectors to dynamically conduct e-discovery within these platforms, as well as involving experts and using repeatable processes.

Additional steps information management (IM) professionals can take to tackle these challenges – and prepare for future e-discovery requests – include:

- Lean on internal knowledge from the IT experts who helped implement the existing systems and applications, so cloud data sources can be incorporated as part of the organization's data map.
- Establish a collaborative dialogue with IT to improve future decisions about cloud providers and apps, giving IM and e-discovery teams the opportunity to examine whether cloud providers have export or data interface capabilities that meet e-discovery needs.

- Work with in-house counsel to ensure new data sources are incorporated into legal hold policies and clearly described in litigation hold notices. Similarly, counsel, IM, IT, and information security teams should align around policies that govern and monitor the security and retention/deletion permissions within cloud collaboration and chat platforms.
- Work with in-house counsel to ensure new data sources are incorporated into legal hold policies and clearly described in litigation hold notices. Similarly, counsel, IM, IT, and information security teams should align around policies that govern and monitor the security and retention/deletion permissions within cloud collaboration and chat platforms.
- Take a holistic approach, collecting information across all business units about the various ways collaboration and chat tools are used and training all users on how to use them within the boundaries of organizational data policies.

Two Cases in Point

The two examples below describe where these challenges came to life.

De-Duping in G Suite

Heading up e-discovery for a client involving more than 88,000 documents from G Suite, the team found that many the documents in the collection – particularly among those in Google's equivalent of Word, Excel, and Power Point formats – appeared to be duplicative or displayed very minor variations from document to document.

The six custodians under discovery had a high volume of data that needed to be reviewed, and the team was seeing false positives across thousands of documents. With a little digging, the team found that G-suite saves documents by making a copy of the data every few seconds, which for the custodians of this matter led to more than 20 versions of each document and more than 60% duplication of unique documents.

Every version saved was collected, and because there were only minor variations among them, the e-discovery software's de-duplication features did not work properly. To deal with this, the team produced only the last current version of each document as of a particular date. Using this method, we reduced the original set by approximately 96% and avoided the rewind review of tens of thousands of duplicative document versions.

Using APIs to Crack Salesforce Collection

Data from Salesforce can be especially critical to e-discovery given the roadmap it provides to sales contacts, internal owners of various relationships, how they are all connected, and other market intelligence that can be relevant to a case. In a government investigation, the team was tasked with collecting from Salesforce under a short timeframe.

The platform does allow export of data, but like many cloud solutions, it can be much slower and more complicated to get information out than it is to import it. Because of this, the team was working to use the "front door" to collect the data and found that it would take more than a month and a half just to complete the collection.

Given the tight deadline for the matter, the team needed a more efficient approach, so it instead used APIs to connect to Salesforce and download the records needed. This was exponentially faster, allowing the team to collect approximately 20 million documents in just eight hours. Ultimately, the team completed the collection and subsequent review by the regulator's deadline.

Three 'Ps' for Solving Problems

Office 365 and other cloud-based solutions are solving a lot of problems, making storage more manageable, and increasing efficiencies, but as outlined in the examples above, the e-discovery "gotchas" are just starting to emerge. An approach that leverages the practical tips that are given above and balances between people, process, and technology can help achieve efficient e-discovery on cloud datasets in the following ways.

The Right People

Any e-discovery matter that involves cloud data should be led by experts with hands-on experience in legal discovery. The team should include professionals with a deep understanding of how to use and manage APIs and extract, transform, load (ETL) processes for database usage and data integration. It is also important that those working with the cloud data are familiar with the matter's key metadata, including which dates, people, organizations, and sources are likely to lead to both relevant and duplicative documents.

The Right Processes

Systems must be set up for API discovery and data profiling, with workflows standardized around these processes. Standard documentation can be put in place to maintain consistency across all matters. Workflows must include a quality testing model for quality assurance and a maintenance protocol to enable teams to replicate workflows across all matters.

The Right Platform

Platforms must have the capability to leverage APIs to ensure versatile, scalable, and secure data integration. It is critical that counsel be able to view the data in a meaningful way; developers may be familiar with extensible markup language (XML) or javascript object notation (JSON), but lawyers see only blocks of text with illegible letters and symbols.

The platform must also allow integration of the cloud data with other e-discovery sources so all evidence can be reviewed holistically. Rapid development of reusable data connectivity components is another important feature that will allow workflows to be standardized across an organization's entire cloud e-discovery portfolio.

Keeping Pace with Technology
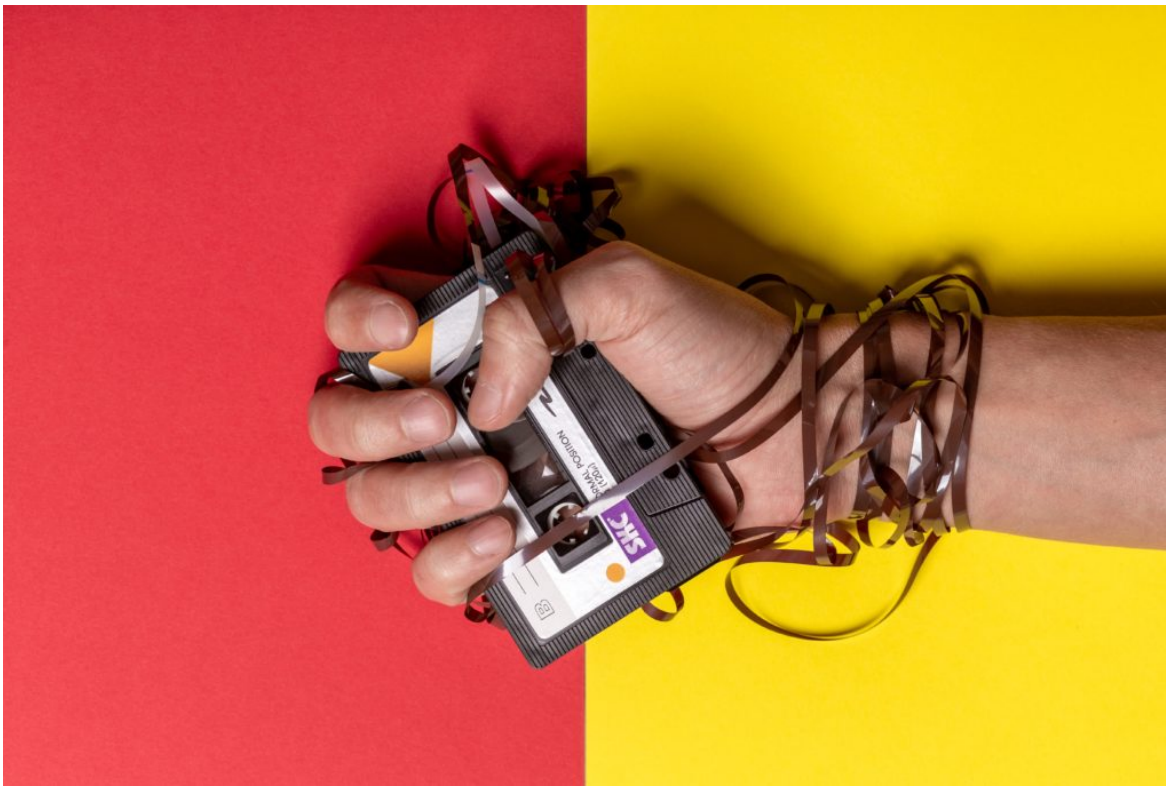
Cloud data is yet another item on the continuum of e-discovery and part of the ongoing struggle attorneys and other e-discovery professionals face to keep practices and workflows apace with evolving technology. As adoption of Office 365 and the emergence of new digital data sources continue to skyrocket, those involved in e-discovery must understand the challenges and be prepared to adjust their standard e-discovery approaches accordingly.

Photos by Stas Knop from Pexels

MORE ON EDISCOVERY
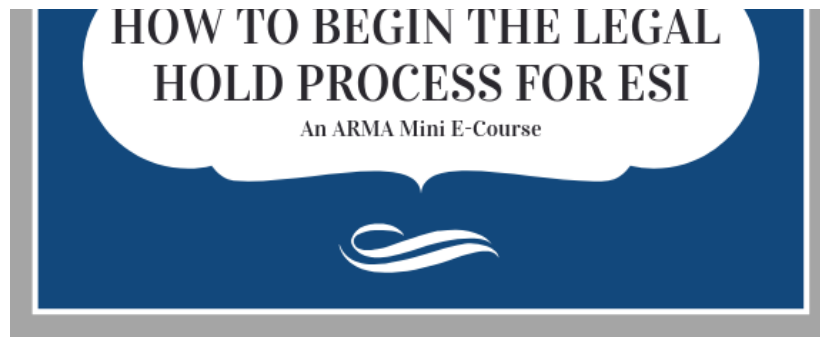
Rapid Resource E-Course: How to Begin the Legal Hold Process for ESI

This first of ARMA's new series of free Rapid Resource Modules takes key content from the popular ARMA book 7 Steps for Legal Holds of ESI and Other Documents, by John J. Isaza, Esq., and John J. Jablonski, Esq. Specifically, the e-learning module covers steps 1 and 2 in the process.

Take The Free Training

MORE ON EDISCOVERY

Checklist for Properly Managing Your Records In An eDiscovery World

Your quick helpful one-page checklist for properly managing your Records in an eDiscovery world.

Download

About the Author

Tim Anderson

Tim Anderson is a managing director in the FTI Technology segment based in San Francisco. He has more than 15 years of legal technology experience as an application development manager, programmer, systems integrator, and consultant. He specializes in developing strategies for preserving, collecting, analyzing, reviewing, and producing electronically stored

information in enterprise data sources, ranging from traditional repositories to cloud-based systems. Anderson can be contacted at Tim.Anderson@FTIConsulting.com.

in

2019.01.22
Stuck in Rewind? Dynamic E-Discovery for Cloud Data

Cloud

Tim Anderson

Tim Anderson is a managing director in the FTI Technology segment based in San Francisco. He has more than 15 years of legal technology experience as an application development manager, programmer, systems integrator, and consultant. He specializes in developing strategies for preserving, collecting, analyzing, reviewing, and producing electronically stored information in enterprise data sources, ranging from traditional repositories to cloud-based systems. Anderson can be contacted at Tim.Anderson@FTIConsulting.com.

Trending Articles

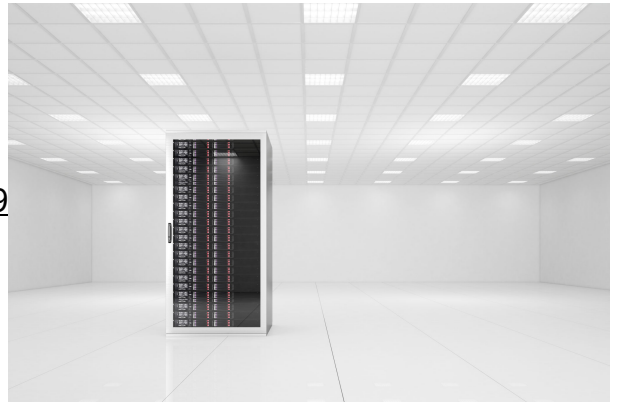[Migrating Legacy Records – a Case Study](#) by [Uta Fox, CRM, FAI](#)

['Alexa, Fix My Records!' A Look at AI in…](#) by [Jeff Hutchings](#)

[Gartner Foresees the End of On-Premises ECMs in 2019](#) by [ARMA International](#)

Big Bucket Retention: Objectives, Issues, Outcomes   by

William Saffady, Ph.D, FAI

Why Information Architecture is Vital to Information...   by

Kevin Parker

Popular Articles

Migrating Legacy Records – a Case Study   by Uta Fox, CRM,

FAI

Recent Articles

[ARMA Q&A: GDPR Regulations and Electronic Records](#)

[Utah to Become Leader in Digital Privacy](#)

['Alexa, Fix My Records!' A Look at AI in the Information Profession](#)

[The Next Internet: Will it Serve the Public's Interest?](#)

[Migrating Legacy Records – a Case Study](#)

Topics

[Analytics](#)

[Archiving](#)

[Artificial Intelligence](#)

[Asia](#)

[Big Data](#)

[Blockchain](#)

[Chapter Resources](#)

[Cloud](#)

[Social Media](#)

[Taxonomy](#)

[Uncategorized](#)

[Vital Records](#)

[Workflow](#)

© 2019 **[ARMA International](#)**