



Litigation Readiness: How an Oil Company Organized and Reduced Their Stored Data

Houston-based Gimmel Group (“Gimmel”) helps corporations and public agencies to better manage their information and records. Brian Tuemmler, a Director at Gimmel’s California offices, leads a special project team that helps clients create and implement processes for filtering accumulated digital information in response to commercial value and legal retention obligations, permitting clients to identify—and defensibly delete—unnecessary electronic files that burden a company’s data repositories without contributing value. Put another way, Gimmel helps clients create order out of the virtual chaos that defines how too many corporations store and manage their electronic information.

The Problem

Tuemmler’s project team was approached by the legal department of a multinational oil company hoping to decrease the amount of network storage it was consuming, while also improving the accessibility of its electronic files. The company had permitted departments and offices to locally manage their hardcopy and electronic files. As a result, even within a single department, users might have to navigate separate (and partially redundant) silos of information. Finding documents created and stored in another location was impossible without bringing in remote help to search and affirmatively send documents to a user. Ultimately, the client hoped to completely retire the shared network drives that its employees had used for years to store files in favor of a centralized web-based solution.



In addition, the legal department realized that to the extent that other departments' records fell under a legal hold, collecting data was a complicated and expensive proposition because data had to be retrieved from individual data silos and even from individual computers within offices. Overall, the legal department believed a different approach would help it significantly reduce its ongoing data management costs while also giving the department better tools to identify and harvest additional value from its existing information archives.

The team worked with the client and its internal records management professionals to identify their content management solution requirements, with a goal of providing a single access point for legal department files, regardless of user location. Though time-consuming, selecting and configuring the new system was actually the easier part of the task. The greater challenge came in sorting through existing files to filter out "e-junk" and to organize retained files so that they could be quickly identified and located in the new system.

The client was particularly concerned that many of its core documents had evolved into informal templates over the course of time, going through countless revisions as they were updated to reflect different business conditions or legal requirements. Dozens of variations on a basic contract might exist, each with unique language that may or may not be relevant in another situation. Ideally, the client wanted its web repository to group documents that had been created from a common source and show each version in the order that it was created. However, finding and organizing all versions of a document revised over many years was easier said than done. Neither the company nor its legal department had required its employees to follow specific versioning or naming conventions, so files with near-identical content might have been saved under completely different names with different authors and with different metadata. To make things even more complicated, some historically significant versions of the files still retained the legacy "8 + 3" naming convention (*i.e.*, "myfile.doc") that communicates almost no information about a file's content. To meet this challenge, the team turned to Equivio.

The Solution

Beginning with an accumulation of hundreds of thousands of employee-generated documents, the team used a variety of tools and processes to winnow down years of accumulated files. The project team worked closely with the legal department to define its business, regulatory, and legal needs. From these requirements, the team developed objective guidelines that it then applied to the document repositories. Even preliminary results were impressive. Approximately 18% of the accumulated files fell into the "junk" category and could be purged from the system. Another 14% of the remaining files fell into defined records retention schedules but had been held beyond their required life, even though they were not subject to litigation hold. Again, these materials did not need to be migrated to the new document repository. A further 9% of the remaining files were older



than seven years and were no longer relevant to ongoing business operations. And 3% of the remaining documents were exact duplicates that could be purged without any loss of business information. Without adding hardware or increasing the client's electricity and cooling bills, the team had increased the legal department's potential network storage capacity by 40%.

After this initial work, however, several hundred thousand documents still had to be imported into the new web repository. And, because of the silo nature of the repositories from which the documents had been drawn, classifying documents by author and previous storage silo failed to link related documents that were spread across multiple locations, much less create the document versioning requested by the client. Enter Equivio.

Equivio near-duplicate identification was applied to the document collection to find commonality among all documents in the file collection. Resulting Equivio EquiSets were then sorted by size. Larger EquiSets (30+ documents) generally indicated that a document had been used as a template. Smaller EquiSets tended to encapsulate the life of a single document over time as it was initially drafted, circulated for comment, and revised for its final version. Using this analysis, the team was able to identify key documents (in the case of templates) while also enabling accurate document versioning within smaller EquiSets as they were imported into the web repository.

The team also applied Equivio to the PST e-mail archives stored on the shared network drives by most members of the legal department. Even though non-duplicate e-mail messages could be imported into the new repository as individual records, a substantial number of the e-mail attachments to these messages were exact duplicates and could be safely deleted. Because the Equivio analysis was run on an incremental basis, the team and its client were also able to begin working with documents almost immediately, without fear of losing or skewing relationships between documents processed at different points in the project. Additional Equivio analysis might increase the size of an EquiSet, but this information supplemented rather than replaced existing analysis already derived from the document collection by Equivio.

In content migration projects, human analysis and data entry is still needed to finish profiling the documents that have been imported into a new repository. In this case, however, thanks to the Equivio EquiSet groupings, every version of a document can be profiled simultaneously, substantially reducing the total effort. In addition, by starting with the largest EquiSets, the legal department will substantially raise the odds that its most important documents are fully profiled early in the process.

Equivio analysis has also provided an indirect benefit to the legal department's business continuity plans. Sampling strongly suggested that documents with multiple iterations contained much higher business value than individual documents that fell outside EquiSets. The project team and the client used this information to rank the likely



importance of documents in its repository, helping implement a tiered disaster recovery solution designed to restore important documents on a higher priority than secondary materials.

Developing artificial document versioning is not a new request for Gimmel. However, past approaches have not always been reliable, since they have relied on date stamp metadata rather than substantive analysis of the documents themselves. "Equivio is the only tool I've seen that can do this version analysis," says Brian Tuemmler, one of Gimmel's Directors. "As for Gimmel," Tuemmler says, simply, "we will be using Equivio to help manage legacy information again."

About Equivio

Equivio enables the management of data redundancy in content-centric business processes. Equivio's technology zooms in on unique data, allowing you to read less, think more, win big™.

With products for grouping near-duplicates, capturing email threads and automating prioritization, Equivio powers a broad range of business applications, including e-discovery, records management, email archiving, data retention and intelligence. To learn more about winning with Equivio, email info@equivio.com or visit www.equivio.com.

Equivio Inc.
5260G Nicholson Lane, Suite 150
Kensington, MD 20895
USA

www.equivio.com