DocAuto

Solutions For Enterprise Content Management

Merging ProLaw into WorkSite Using DocAuto Tools White Paper

October 2011

INTRODUCTION

McAfee & Taft, the largest law firm in Oklahoma, acquired boutique litigation firm Eldridge Cooper Steichen & Leach ("ECSL") in the fall of 2011, and needed to merge ECSL's Thomson Reuters ProLaw® document management system into McAfee & Taft's existing Autonomy iManage WorkSite® document management system. The merger of these systems involved several million documents, some of which were directly managed by ProLaw, some of which were not, preserving the ProLaw organizational structures and providing for a rapid turn-around over a weekend as the merger was consummated. The entire project had to be planned and executed in less than two months.

DocAuto's iImport Content Importer for WorkSite application was used to perform content importing; updating metadata to reflect new metadata values inherent in this type of conversion process; preserving imported document information to eliminate redundant importing or to perform differential updating or version importing if necessary; and to create "placeholder" documents in WorkSite to reflect "Event" records in ProLaw that had no physical document behind them, but relayed important information to users.

DocAuto's WorkSpace Manager application was used to recreate identical WorkSpace structures that corresponded to both the physical structures of the documents on the back-end of the ProLaw repository, but also user-created *ad hoc* structures that were represented within the ProLaw interface and stored in the database, not necessarily captured as physical structures on the back-end file systems. WorkSpace Manager also added documents to one or more folders where they may have existed in the corresponding ProLaw structures to preserver the content-to-structures linkages. DocAuto refers to this process as "Provisioning".

WorkSpace Manager also "Refiled" the profile metadata from the structures to the documents as a way of verifying that all documents had been stored in the correct Client/Matter WorkSpaces and Folders. This process also made it easy to identify documents that had ambiguous metadata because of being placed in multiple different Client/Matter WorkSpaces, or documents that for one reason or another may have failed the Provisioning process. Ultimately, only documents that had been imported from "raw" folder shares that had not been managed by ProLaw required manual correction to one degree or another.

David Kiefer, President of DocAuto, spearheaded the DocAuto team for this project, and Liz Groom, Director of Information Systems at McAfee & Taft, worked around-the-clock during the project to perform the vast bulk of the hands-on work that made this project a success.

SPECIFIC BUSINESS CHALLENGES

ECSL needed to continue business as usual until the last possible minute of the switchover of the systems. The plan was that an early snapshot of their ProLaw documents and data would be used to drive an initial conversion pass, then differential snapshots would be used to gather differential information to perform one or more updates up until the last minute, when their staff would start training. When they started working in their new iManage environment, everything would be reproduced in identical structures (in WorkSite rather than ProLaw), creating a seamless transition and an immediately productive environment.

Another issue was the fact that the metadata used at ECSL was of course not used in the same manner at McAfee & Taft, so a data conversion or "translation" would be necessary for User IDs, Client numbers, Matter numbers and other identifying information. Closed matters were archived directly, and allowed to retain their "legacy"

identifiers, but equivalency tables were set up in the accounting system to convert old values to new values on-the-fly.

Specific Technical Challenges

Several challenges were obvious at the outset of the project, others arose as the project was underway that had not been anticipated, but through careful analysis of the source data, adjustments were possible to address specific resource, content or data issues.

ProLaw Database Structure

The largest problem was determining *how* ProLaw was storing the majority of its data. Some of the different approaches applied for understanding how structural data was being stored and extracting that information proved incorrect, but finally a solution was determined that identified correct, efficient queries and other processes without difficulty. Additionally, there were issues associated with data integrity (eliminating circular references that had managed to be created in the ProLaw data) that were unpredictable and unexpected, but the unusual data was identified and this was cleaned up manually. In general, strange or corrupt data accounted for only a small fraction of one percent of the total data within ProLaw.

Non-Document Items in ProLaw

ProLaw has the ability to store "Note" items in its database that correspond to reminders, notes or other types of annotations within its system. While these items have records in the ProLaw database, they do not correspond to physical documents in the ProLaw document store. This represented an unusual challenge, because the firm wanted these items to be preserved and displayed in the same context within WorkSite as they did in ProLaw.

DocAuto was able to achieve the desired result using iImport by having a simple job that cycled through all of the Note records in the ProLaw system, and created "placeholder" documents in WorkSite that would correspond to the original ProLaw Notes. These documents had minimal content that indicated to the user that these documents were placeholders, but all of the original data from the ProLaw Note was converted to the new WorkSite document's profile.

Preserving Document Links to All Locations

Like WorkSite, ProLaw has the ability to link a document to multiple structures or folders within its interface. DocAuto's WorkSpace Manager application was able to preserve this functionality using its Provisioning capabilities, which allow any document to be linked to any number of folders in any number of WorkSpaces. WorkSpace Manager also allows for documents to be referenced across multiple WorkSite databases using Provisioning, or to be "migrated" from one database to another using Provisioning or Aging, but these operations were not necessary to achieve the desired results at McAfee & Taft.

Importing Performance

McAfee & Taft's conversion timetable was extremely aggressive, and a large volume of content and data needed to be converted as part of the initial pass, then subsequent differential passes were used to bring in more and more current updated content and new content, down to the final conversion pass. Because of the large volume of content, it was necessary to scale up the conversion process using iImport to achieve maximum throughput. The core limiting factor within these operations is generally the performance capabilities of the WorkSite SQL

server, which must handle the large volume of transactions that are being fed into the system through the WorkSite server by the DocAuto three-tier applications.

Through the use of multiple machines to run importing jobs in particular, and minor adjustments to various "accessory" databases that were being utilized to hold "translation" data that was used to convert metadata during the import, and to track the progress of imports to perform differential processing in subsequent passes, the team was able to achieve maximum throughput that allowed tens of thousands of documents to be imported per hour, and tens of thousands of "structural" operations, such as creating WorkSpaces or Folders within WorkSpaces, or adding documents to folders, to be performed per hour. Total "maximum load" processing of the conversion of several terabytes of documents took less than 300 hours.

Conversion Methodology

As with many corporate mergers, little was known about the needs of the different user communities until the last minute. This of course dictated a tight timetable, and little leeway to make adjustments based on any unexpected technical issues. It was important then to develop a simple "plan of attack" that could be adjusted agilely as the project progressed, and could be adapted to changing needs and unforeseen obstacles, both due to technical issues or user requirements.

Importing Documents

One of the most straightforward parts of the conversion was defining the document importing process. Since the introduction of DocAuto's iImport in 2003, and with additional enhancements since, it was a logical out-of-the-box solution for this process.

Metadata Changes

The ProLaw metadata did not correspond to the metadata used in the WorkSite system. A simple "translation" database was set up that allowed a simple query to run on a file-by-file basis to get the "translated" data needed to populate the WorkSite profile with appropriate values that corresponded to the original values in ProLaw.

Active and Archive content was also treated differently – active content was to have new client, matter and other identifier values assigned that corresponded to the active billable matters in McAfee & Taft's Elite system. Archive content was to be flagged with a single "archived" client and matter metadata combination, but the "original" or "legacy" firm client/matter information was to be preserved on other fields on the documents' profiles. This would allow users to find content associated with the legacy closed matters in the future.

Maximizing Throughput Using Multiple Importing Processes

A vital measurement to the success of the project, as mentioned earlier, was speed and maximum throughput. As iImport is able to process any number of sources within a job, and DocAuto's licensing scheme does not limit the number of physical machines or instances that can be run simultaneously, it was fairly easy to scale-up processing and throughput capabilities by cloning virtual machines that had been configured for this purpose. At the maximum processing load, about a dozen machines could be running simultaneously, which would keep SQL utilization at around 90 percent, which was deemed to be the safe maximum level. As noted above, SQL performance is generally the limiting factor – it was not necessary to use other instances of the WorkSite Server other than the production system. WorkSite Server performance peaked at around 2,000 – 3,000 transactions per second.

Tracking Import Results

As differential importing was going to be a major aspect of the merger and conversion process, as well as importing documents that had multiple versions, and preserving the version set within the new "destination" document in WorkSite, an accessory database was set up to allow iImport to store document-by-document processing information. This data could then be used in subsequent passes to determine if a document had been imported already, or whether the document being processed was an edited version, a new version or a new document that did not exist in a prior pass. Again, iImport had this functionality already for several years, so it was a simple matter to create a suitable database for this information. Adding and optimizing SQL indexes on these tables and columns within this and other accessory databases was critical to optimizing the performance and throughput of the process.

Differential/Update Importing

Because iImport had preserved detailed information about each individual file as it was processed in the accessory database, it was easy to use iImport' s "branching" capabilities during import to control the import flow on a file-by-file basis. Files that had been imported previously and had not been edited could be skipped. Files that had been imported previously but had been edited could be overwritten in the WorkSite system, or saved as new versions. Documents that had been created and had not been previously imported could be imported and their data captured for subsequent passes. New versions that had been created in ProLaw could be imported as corresponding versions of the same document in WorkSite.

Creating Documents Where Documents Did Not Exist

A unique use case arose to create documents in WorkSite that corresponded to the Note items that were stored in ProLaw. These Notes do not actually have a physical file in the ProLaw document store, but their metadata as stored in the ProLaw database needed to be preserved and displayed to the user in the same location in WorkSite. A simple looping job was configured in iImport that would take these Note records one at a time and import a placeholder file in WorkSite, populate the WorkSite profile with the same metadata from the ProLaw Note, then write back to ProLaw that the Note had been processed, so that it could be skipped on the next pass. While a somewhat unusual use case, iImport was able to handle this task out-of-the-box.

Creating Structures in WorkSite

The process of creating matching structures in WorkSite to correspond to the structures '*as users saw them'* in ProLaw presented several challenges:

1. Only the lowest level or root structures in the ProLaw interface actually corresponded to physical locations for the content in the ProLaw document stores.
2. Users had the ability to extend the basic root structures to create their own subfolders, sometimes many levels deep. (The deepest level we encountered was about twelve folders deep.)
3. Documents can be associated with multiple locations in the ProLaw system. This presented challenges in determining what metadata should be applied to the document based on where it was "filed".

The solution involved using WorkSpace Manager to use a combination of data from the structures of the physical file stores, as well as data from the ProLaw database. Generic WorkSpace templates were created that could recreate any folder structure as it had been customized by the users over the year in ProLaw. Utilizing this process, test runs were performed to recreate particularly complex or deep folder structures, which were then checked against the originals in ProLaw. While WorkSpace Manager's performance could have been significantly

increased by adding multiple instances of the application running simultaneously, the standard throughput of over tens of thousands of folders per hour was determined to be adequate for this project.

Importing Documents into (Multiple) Folders

The folders that were created by WorkSpace Manager were flagged with unique identifiers that corresponded to the source ProLaw folders. This data was then used to "re-link" the documents after they were imported to the same combinations of structures in WorkSite. This was done using the native Provisioning functionality of WorkSpace Manager, which can identify content to be added to a folder using "Optimized Provisioning" jobs that have been documented in the WorkSpace Manager Job Strategies Guide, which is available of the DocAuto.com Website.

Refiling Metadata and Security

The final step of the process was to "refile" the security and metadata of the WorkSpaces to the folders and documents after they had been imported into WorkSite using iImport and added to the correct WorkSpaces and folders using WorkSpace Manager. Developed in 2006 in cooperation with Interwoven (at the time) and the world's largest user of WorkSite, DocAuto's Refiling Server Enterprise was used to perform this operation. Refiling Server has unique built-in capabilities to deal with documents that are stored in multiple folders, and can combine or "accumulate" security from all of the multiple folders where a document may reside to assure that the most or least restrictive security is honored.

About DocAuto

DocAuto is an internationally utilized solutions and services provider focused on Enterprise Content Management and E-mail Management for iManage Work and SharePoint. DocAuto prides itself on rigidly working within iManage Work's 3-tier architecture, Microsoft's SharePoint architecture, and by following Enterprise Content Management best practices. At DocAuto we are committed to creating innovative solutions that improve our clients' unique work environments, foster growth, and meet the future needs of our customers.

Telephone US Toll Free: (800) DocAuto
Telephone UK: 020 3582 5070
International: (770) 242-6747
Web: www.DocAuto.com
E-mail: Sales@DocAuto.com