

The WebDAV Decision

Author: Gary M. Gershon
Principal, Intermedia Sciences
28 Branchville Road
Ridgefield, CT 06877
203-431-9328
gershon@celsus.net

Abstract

This article describes the decision process used at Chubb and Son, a well-known property and casualty insurer and financial services provider, to move from document management solutions that relied on proprietary programming interfaces, to the use of WebDAV, a vendor-neutral Internet standard protocol for accessing and managing documents. This strategy allowed their line-of-business applications to develop thin- and thick-client application code to interoperate with back-end document repositories from various vendors and to share documents among business areas that were previously islands.

Chubb workstation users can also avail themselves of the built-in WebDAV client support provided in current versions of Microsoft Office as well as in Adobe Acrobat. This innovative corporate solution was implemented using the WebDAV-compliant Web File Server product from Xythos Software, Inc., which provided enterprise-wide scalability, server-side programming interfaces to customize security and integrate with the business applications, as well as other required functional and document management capabilities.

Chubb is enjoying both short-term savings compared to the proprietary offerings from most traditional document management vendors, as well as long-term savings by allowing servers to be replaced and upgraded without reprogramming the client application code. The new ability for a common workstation client to directly access documents from multiple departments will speed business processes, improve decision-making, and enhance CRM efforts.

Introduction and Background

The Internet has created a new impetus for companies to reengineer their corporate document management solutions, as they move to make content and workflows available to users via the Web. Thin-client initiatives are gaining momentum to provide Web access for extranet use, as well as to support work-at-home and mobile employees. Business process owners are striving to improve efficiencies by receiving transactional documents from partners and customers electronically—and then processing and editing them via Web-based workflows. These were significant drivers for Chubb and Son to initiate a review of its legacy document management systems and needs last year. The result was a decision to utilize standards-based solutions whenever possible, replacing, over time, their proprietary document management systems.

Chubb, like other substantial insurance and financial services companies, has moved significant portions of its paper files into electronic format as a means of reducing expense and improving customer service and efficiency. Traditionally, the choice of vendors was made independently by each line of business and department, to best meet their respective needs. Thus commercial lines used a different solution from personal lines, as did claims, general counsel, and other financial services departments.

Chubb undertook a cross-departmental study in the summer of 2000 to determine what were the outstanding requirements for document management—both thin- and thick-client—and to ascertain how these needs might best be addressed. This study led to a focus on the use of Internet standards for document management, a pilot, and the selection of the WebDAV-compliant Web File Server product from Xythos Software.

The result was a new approach to implementing document solutions at Chubb that offered substantial long-term savings, and which was sufficiently low in cost that implementation could be begun immediately.

Requirements Study

The largest five application areas at Chubb participated in the document study, led by a senior experienced consultant. Each area provided support to review and document their requirements under a number of categories, including:

The WebDAV Decision

- Document capture,
- Repository functions,
- Indexing,
- Web enablement (browser; thin-client),
- Workstation Access (Windows; thick-client),
- Access control and security,
- Workflow,
- Searching, and,
- Document retention.

The result was a comprehensive matrix that allowed, for perhaps the first time, the collective requirements of these key organizational units to be analyzed as a group. During the process, the needs of smaller departments were factored in under the philosophy that by finding a solution for the largest departments then the small ones would, in most cases, be able to coattail on this common solution at minimal additional expense.

A number of interesting conclusions were drawn from the study and the resulting consolidated matrix, which may be valid for other organizations:

1. The need to effectively support thin-client document access had become universal. The existing deployed implementations had been built for thick-client use and did not provide for HTTP access, nor were they able to meet web scalability requirements without making substantial new investments.
2. An affordable vendor solution was needed for storing documents that would provide more functionality than was offered by the existing implementations, some of which had been developed in-house using SQL tables. For example, versioning, foldering, and improved document recovery facilities were needed to protect and organize growing document collections.
3. For the large departments, there was no significant need for vendor-supplied user interfaces—thick or thin. These organizations typically had thousands of users and usually created a custom interface or GUI, commonly referred to as a “workstation,” which maximized productivity by supporting workflow processes, unifying information from separate back-end systems, and incorporating business-rule logic.

The WebDAV Decision

4. The same was true for vendor-supplied indexing and search facilities. These departments utilized their own SQL tables and indexing and search facilities which were fully integrated with the rest of the application. A single mouse click was often sufficient to view documents related to a client from a client-information window, without needing to re-key client or policy numbers.
5. Capabilities were required to enable documents to be displayed and edited using various standard vendor client applications. In particular, Microsoft Office and Adobe Acrobat needed to be supported for both thick- and thin-clients.
6. Due to the tight integration with the business applications workstations, it was necessary to be able to implement the document management capabilities by application programmers developing with Smalltalk, Java, and Basic. Workstations were now uniformly on Windows-based platforms, but servers utilized both NT and AIX environments.

The sum of these requirements was that a solution would best take the form of integrated “components”—rather than the typical “soup-to-nuts” solutions provided by traditional document management vendors. These comprehensive vendor packages often included many expensive features that were not necessarily needed, and thus make the cost of implementation—both licensing and integration—prohibitive. For Chubb, the key components would be client-side scanning, indexing, viewing/editing, and workstation integration components—most of which were already in-place, and server-side repository and Web-enablement capabilities. The new ingredient was that Internet standard protocols—the DAV family—were now available to standardize the integration of the client-side and server-side components.

The Decision to Adopt WebDAV

Beyond the operational costs of a document management system, the cost of customization and integration programming is the most substantial investment for the various departments. Programming also represents a substantial cost for the subsequent system and platform migrations that inevitably occur. What would be very attractive would be a standard document management programming interface, akin to what SQL provides for databases, which would be vendor, operating system, and platform independent.

The WebDAV Decision

Discussions of document management on the web eventually turn to WebDAV (Web Distributed Authoring and Versioning) which is the Internet standard for managing documents. WebDAV has been widely implemented for Internet collaborative and document sharing applications. Significantly, WebDAV is well supported by Microsoft for both clients (Microsoft Office 2000 and XP, Windows Explorer, and Internet Explorer) and servers (Microsoft Server 2000, IIS, and SharePoint), and thus has gained wide acceptance—due in no small part to it having been fully embraced by this major industry player.

WebDAV is the base standard of the DAV group of document authoring standards, and was formally approved in early 1999. Subsequently, associated document management standards have been generated for searching and locating documents (DASL), for maintaining document versions (DeltaV), and for controlling access to documents (ACL). Information on all of these standards, as well as associated initiatives and products, is available on the web at <http://www.webdav.org>.

The DAV family of standards defines extensions to the HTTP Internet protocol used by Web browsers—and many other applications—and utilizes XML to encode data for transmission. These standards define protocols—essentially the format and sequence of what is transmitted back and forth from client to server. Protocols are inherently vendor, operating system, programming language, and platform independent, yet provide programmers with the means to interact with diverse applications. Programmers will often utilize lightweight objects or classes to encapsulate and simplify their programming effort. Several libraries of such code are available for Java as open-source code, and Microsoft provides comparable functionality via ActiveX controls. Protocols based on HTTP and XML work perfectly well between different platforms, so that, for example, a Visual Basic client can issue WebDAV requests to either an NT or a UNIX server and receive comparable responses. This environment provides a great deal of flexibility and portability compared to the traditional proprietary APIs (Application Programming Interfaces) that were provided for specific programming environments.

The WebDAV standard itself defines the basic set of capabilities needed to work with documents—store, retrieve, copy, move, rename, delete, lock, and unlock. Additionally indexing information can be stored as DAV document properties, which can be retrieved and

The WebDAV Decision

updated by WebDAV and effectively searched using DASL. Properties are simply keyword-value pairs, such as *author=JM Smith*, *client=76436827*, or *retention=2004-07-16*.

Observing the widespread and rapid implementation of the DAV standards by Microsoft, Adobe, and other industry leaders, the Chubb team recognized that implementing a document management solution that incorporated WebDAV would provide a number of short-term and long-term benefits and cost savings.

In the short-term, the major benefit was being able to adopt the Internet model of assembling solutions. The Internet has been built on a component and standards-based model since its inception. Key standards, such as HTML, HTTP, TCP/IP, and increasingly XML, have allowed organizations to buy and assemble components, such as authoring and development tools, web servers, app servers, and client browsers, from a variety of vendors with confidence that they will work together—albeit sometimes with some minor adjustments.

With WebDAV and a few de facto content standards, such as HTML, XML, TIFF, PDF, and Microsoft Office, this component-assembly model can be applied to document management. As long as the documents can be sourced in a standardized format, they can be managed on a server supporting WebDAV, and viewed and edited using one of the appropriate viewers from both thick- and thin-clients. Should one of the initial component choices need to be changed in future years, it can be re-sourced without necessarily affecting the other component choices.

Thus Chubb, by selecting a standards-based approach, could avoid having to invest the time and expense required to identify a single document management vendor. In the long-term WebDAV would offer a means for Chubb to avoid the substantial costs and effort required for migrating and testing their investments in application programming code as document management systems were replaced. Chubb would not be encumbered by the high switching costs and lock-in associated with many vendor solutions—as well as in-house developed repositories—due to proprietary APIs and content formats.

The result of these deliberations was a decision to see if the departmental requirements could be met in a cost effective manner using components and programming based on WebDAV protocols and other standards, rather than attempt to conduct a RFP-like process for vendor

selection. Chubb would be joining the trend to assemble its solution from interchangeable standards-based components—a shift that represents a sea change in the way the document management industry has operated for the past two decades.

Solution Alternatives and Vendor Selection

The core requirement that needed to be met in this approach was to find a WebDAV-compliant document repository. Other needs, such as scanning or workflow support, could be met using well-known standard industry component packages that would utilize WebDAV capabilities to store and retrieve documents. WebDAV relies on standard HTTP URLs to reference documents, so that workflow needs could be expected to be readily satisfied using either collaborative email approaches, or more structured SQL-based routing engines.

The repository would need to meet a number of requirements, some of the key ones being:

- It must be scalable to enterprise-wide levels to handle thousands (or tens-of-thousands) of users.
- It must provide an integrated user and document security capability, but must also be readily integratable with existing, and future, Chubb authentication and authorization systems.
- It must provide the features that were required by the departments, including versioning, foldering, recoverability, and activity logging.
- It needed to provide a rich and full implementation of WebDAV, with the ability to enforce document-related business rules when processing WebDAV requests and events.

There are several dozen WebDAV server implementations available, each targeted to a somewhat different business or application need. The effort stayed focused on the requirements and a product offered by Xythos Software rose to the top of the list. Xythos' product met all the requirements, fit well into Chubb's infrastructure, and was relatively quite affordable.

Another choice would have been to select an offering from the ranks of the traditional document management vendors. These vendors, however, have been conspicuously late-to-market with their WebDAV offerings. Perhaps this is due to engineering problems in fitting standards-based Internet capabilities onto existing products, or this may be due to marketing and business decisions.

The WebDAV Decision

Lastly, had Chubb wanted to construct its own WebDAV server solution, there are certainly ample programming components—most as open source—as well as complete documentation on the protocols available. However, this would have meant diverting development talent away from solving business problems in order to develop and support software that was already available at affordable costs in the market. The vendors also have the expertise and resources to participate in ongoing standards-making groups and can provide progressive new releases with improved functionality. This was a traditional multi-faceted "build-and-maintain versus buy" decision.

For other companies and applications, there is likely to be found a WebDAV solution that meets their needs. In particular, there are a number of vendors with departmental-scale solutions available, some tailored to the collaborative and workgroup needs of particular businesses and processes. At Chubb, the approach was to solve the enterprise level needs first, and then deploy the same solution to smaller departments as appropriate. This doesn't prohibit Chubb departments from using a different vendor's or locally developed solution, but it does permit leveraging the greater resources and budgets of large departments to cover the initial investment in infrastructure and education associated with any new technology.

Following an evaluation pilot, the repository product selected was the Xythos Web File Server (WFS). This is a WebDAV implementation, built using Java Servlets and JSPs, which can be deployed on any of the Chubb NT or AIX servers. It can utilize an NTFS or NFS (UNIX) file system to store the documents with flexibility to support standard server attached storage, as well as Storage Attached Network (SAN) and Network Attached Storage (NAS) configurations. Since Chubb will be dealing with terabytes of storage, it is particularly convenient that a mix of these three storage configurations can be employed for a single application, thus simplifying long-term growth.

The Xythos WFS—like the traditional solutions from the enterprise-scalable document management vendors—uses an SQL database to maintain information about the documents, their status, and their locations, thereby offering high performance and scalability.

Over time, it is possible, if not likely, that other WebDAV-compliant servers will be utilized at Chubb. WebDAV can support many types of applications, including configuration management

The WebDAV Decision

for development groups, and maintaining web content and programming. In any case, the corporate investment in application programming and clients is well protected by the use of a standard protocol, rather than vendor APIs, to access the documents.

Implementation

Implementation of WebDAV required installation of a standard Web server/app server configuration, allocation of SQL databases and document storage space, and application-related programming and integration to be implemented at both the clients and servers.

The WebDAV clients at Chubb consist primarily of out-of-the-box vendor software, such as Microsoft Office, and in-house developed business applications. The business applications are written either to run on departmental workstations, or to run on application servers for thin-clients. WebDAV programming components are available in essentially all current languages, and were utilized directly for Chubb's Java and Visual Basic needs. Chubb has a substantial amount of legacy workstation and server code written in VisualAge Smalltalk, and WebDAV programming was developed using standard Smalltalk HTTP and XML facilities to provide support for this environment. With these applications, Chubb was able to replace proprietary APIs with standards-based object-oriented interfaces to store and retrieve documents. The Java clients used open source Java classes, and Visual Basic clients used Microsoft's HTTP/XML ActiveX controls for WebDAV.

The Xythos WFS provides Java source and sample code for the key integration points: authentication, authorization, and client-request event processing. The changes generally involved integrating and testing pre-existing Chubb Java classes into the WFS code at well-defined points.

For authentication, the server Java code was extended to use Chubb's enterprise mainframe RACF system to validate user ids and passwords. Additionally, document access authorization was modified to use mainframe IMS-based systems, which were accessible using IBM's MQ Series messaging, as well as DB2 tables, which defined user roles and associated document-access business rules.

The WebDAV Decision

More subtle programming changes were introduced in the WebDAV request processing flow. For example, some business applications prohibit users from renaming or deleting system-managed documents using out-of-the-box WebDAV clients, such as Microsoft NT Explorer, since the business applications manage these functions. Because the server knows which type of software client is making a request, it was a relatively straightforward task to "lock-down" functions that certain clients were to be prohibited from using. The client software itself, such as Microsoft Word, did not need to be changed. The amount of programming varied by department since the business rules being implemented were specific to these departmental applications.

Migration programming was also developed to move documents from pre-existing repositories to the Xyθος WFS. This was done with a combination of batch mass-migration code, as well as a "on demand" migration approach whereby documents were retrieved from an old repository at the time the WebDAV request was received by the server. This effectively fully web-enabled the old repository on "day one" with all documents becoming accessible via HTTP prior to the mass migration commencing.

Some departmental uses did not warrant particularly high degrees of customization, and for these users the Xyθος WFS thin-client interface (HTML) was used to augment the vendor WebDAV clients with additional functions, such as viewing earlier document versions, or inspecting document history logs.

Future Plans

By utilizing WebDAV as a common protocol to access the documents in each department, it opens the possibilities for sharing infrastructure and customized code, as well as building cross-departmental document sharing and workflow applications. Several initiatives are under way to solve additional needs.

A common requirement among all the departments at Chubb is compliance with corporate records retention rules. Documents must be retained for specific periods, and then either destroyed, or moved to long-term archival storage media. WebDAV offers a new opportunity to easily interface a common records management engine to multiple departmental repositories. By setting WebDAV properties, such as document class and retention action date, consistently in the organization, it is no longer necessary to build a custom retention processing application for

The WebDAV Decision

every department—a single application can examine the properties, delete expired documents, and migrate others to archives.

As the WebDAV servers begin to receive documents directly from business partners and customers, it will be important to provide appropriate gateway processing capabilities. The Xythos WFS product already provides an extranet capable and customizable security and session model. Chubb will want to implement additional capabilities for virus checking and quarantining, and email workflow-initiation messages. These should fit well within the capabilities of WebDAV, assisted by the Xythos WFS event model, to initiate processing.

Departments that were previously unable to readily share documents due to differences in vendor client APIs, can now access documents from other groups, subject to access permissions and business rules. This concept is increasingly important to support CRM initiatives where customer service representatives need to respond to phone requests for diverse information. The participating units can also now share programming code, training, and technical staff more readily since the code and skills are standards-based.

Chubb's IT organization, by focusing on standards-based document management, will be able to provide more capabilities to their business users, better utilize IT talent and resources, and more effectively support CRM and workflow enhancement requests. Their decision to use WebDAV, along with the selection of an enterprise WebDAV product from Xythos, has essentially eliminated the need for their business application programmers to work with proprietary APIs. The result is the opportunity to deliver more to their users at dramatically reduced costs.