

sas | The Power to Know. **sas**

Creating a Home-Grown Document Management System
 Dick Vohlers
 SAS Institute

Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

Introduction

- SAS Institute is the largest privately-held software company in the world, with revenues last year over \$1.5 billion (<http://www.sas.com>)
- My role: Principal Software Developer in charge of developing and supporting the corporate Document Management System (DMS)
- The DMS is an internal system used for storing and retrieving company documents, and is based on Adobe Acrobat (started with Acrobat 5, currently on Acrobat 7) running under Windows

Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

DMS Versions

- The DMS is an umbrella term for several similar though separate versions:
 - COP DMS (Contract Operations—Software Contracts)
 - SCA DMS (Service Contracts Administration)
 - LGL DMS (Legal—where SAS is buyer, not vendor)
 - HR DMS (Human Resources—personnel records)
 - CIP (Customer Information Portal—exposes contracts to customers)
 - In progress: versions for tax records, insurance records, and European contracts

Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

DMS Components

- Each consists of several related applications that are used to:
 - Create & manage the documents (implemented using plug-ins created with the Acrobat SDK)
 - Audit scanning operations & implement retention rules (implemented using VB programs created with IAC)
 - Retrieve documents via a Web Interface on the company intranet (implemented with a Java program—COP & SCA only)
 - Converting PDF pages to JPEGs for display on the Web (CIP only—more later)
 - Update indexing data (implemented with SAS code)

Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

Version Differences—Each version has:

- Its own database (stored in SAS), its own image storage area, and indexing information, some of which is pulled from external sources
- Its own retention rules
- Its own way of limiting access, either through Windows permissions, SSL log-ins, or just limiting access to the necessary plug-ins
- Its own audit trail
- Its own set of PDF bookmarks, and in some cases “child” documents

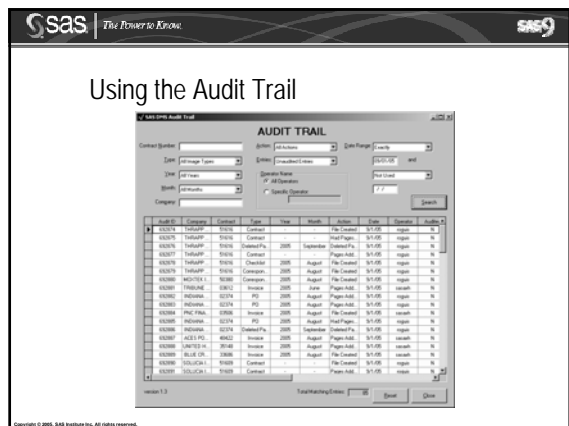
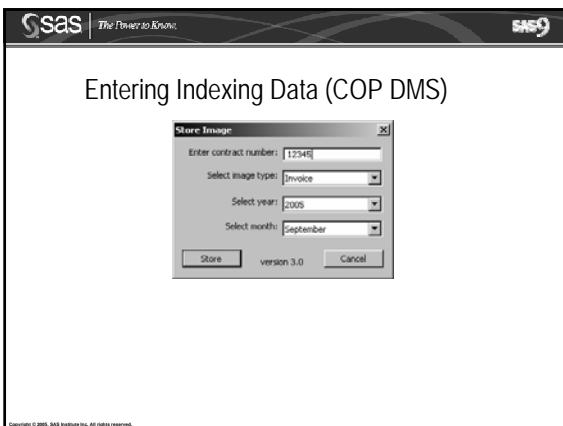
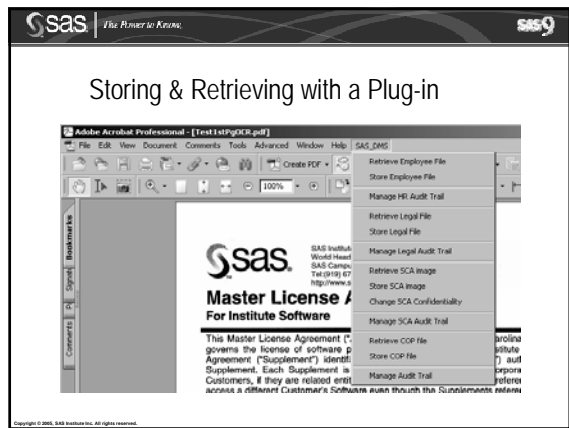
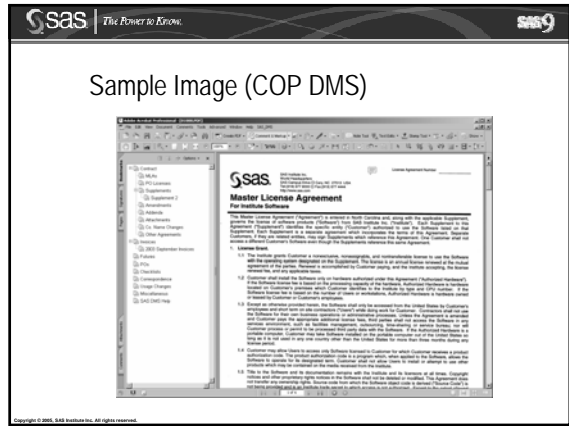
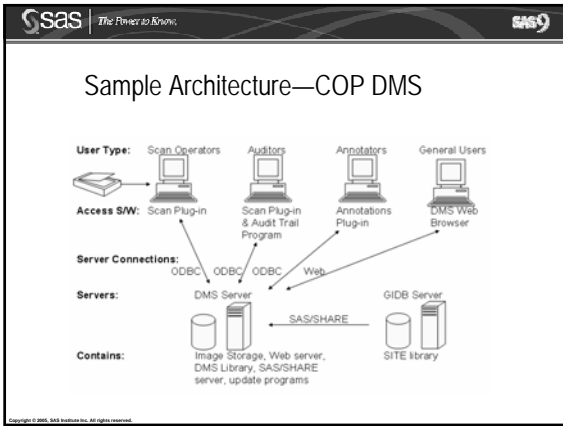
Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

Why Custom Tools Based on PDFs?

- Why custom:
 - Expense of other tools
 - Can implement own retention rules (deletion of files)
 - Ability to have dedicated support
- Why PDF:
 - Standard format
 - Nothing else needed for Web (BUT: CIP problem)
 - Don't have to recreate document management tools
 - Size of files (COP DMS has over 150,000 files)
 - Most users can get by with free Adobe Reader

Copyright © 2005, SAS Institute Inc. All rights reserved.



sas | The Power to Know. **sas**

Code Samples

- Demo of sample plug-in code
- Demo of sample IAC code
- Demo of SAS datasets and sample SAS code

Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

Lessons Learned—Usage

- Scanning is memory intensive—scan machines need at least 1 GB of RAM
- Scanning too fast can cause problems
 - “Open cross-document links in same window” and applying passwords
 - Double clicking a button can skip next dialog
- “Solidifies” processes so harder to change
- Acrobat 6 users can only get rid of “new version” message for one session

Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

Lessons Learned—Programming

- Documentation
 - Hard for a newbie to start
 - Has gotten better with each release, but still has holes or hasn't caught up with changes
 - Need to look in header files or similar commands for other documents
- Snippet Runner can be useful
- AppWizard can be useful, especially for test cases

Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

Lessons Learned—Acrobat Upgrades

- APIs not sufficiently tested or changed before new versions of Acrobat shipped
- Issue with automatic application of password using PDDocAuthorize—worked in Acrobat 5, but not 6
- Upgrade forced on retention app when at Acrobat 5 level; couldn't handle Acrobat 6 image
- Acrobat 6.0.3 update caused scan problems
- Method of calling Kodak Scanner GUI changed from Acrobat 5 to 6 requiring extra button press

Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

Lessons Learned—Undocumented Changes

- Big issue with the Customer Information Portal:
 - Files won't automatically appear on Reader if user accessing using IE, via SSL, and “Do not save encrypted pages to disk” is checked
 - Requires converting each page to JPEGs so can wrap in HTML and present
 - Acrobat's method of saving as JPEGs changed from Acrobat 5 to 6 and from 6 to 7
 - Due to number of documents & pages, forced in Acrobat 6 to use another tool
 - Other tools did not perform well with upgrade to 7
 - Program currently offline until a means to create small readable images quickly via code can be found

Copyright © 2005, SAS Institute Inc. All rights reserved.

sas | The Power to Know. **sas**

Lessons Learned—Getting Solutions

- Online Resources:
 - PDFZone
 - Planet PDF
 - Developer's list at PDF Zone
 - Adobe's User to User Forums
 - Googling
- Get in on testing—that's how I solved the PDDocAuthorize issue for Acrobat 7
- Paying for incidents can (sometimes) help—but don't be afraid to ask for a refund if it doesn't

Copyright © 2005, SAS Institute Inc. All rights reserved.

Summary

- While a lot of work to create and to upgrade with new versions of Acrobat, the DMS is not too hard to maintain on a daily basis, and has been well-received inside of SAS.
- John Boswell, SAS's Vice President and General Counsel (corporate sponsor for the DMS) says "The DMS system has created an incredible amount of efficiency for our lawyers and contract administrators. It also reduced our need for people to pull and deliver files manually."