Battelle
The Business of Innovation

Harvesting Records from SharePoint Sites

Charlie Arp

arpc@battelle.org

(614) 424-7897

Agenda

- Battelle Records Management Office
- Why harvest SharePoint content
- SharePoint - defined and history
- Harvesting content from SharePoint sites
- Problems
- What's next - SharePoint 2010 and ECM
- Conclusions

What we do

- Charitable trust – 501(c) 3
- Scientific research and development
- Reduction to practice and licensing of inventions
- Global science and technology enterprise that explores emerging areas of science
- Manage laboratories for customers

What We Are

- Global enterprise
  - Applying science and technology to real-world problems
  - Managing machinery of scientific discovery and innovation
  - Creating commercial value by bringing new technologies to international marketplace
- Non-profit, charitable trust formed by Will of Gordon Battelle in 1925

- Generates $5.6 billion annually in global R&D
- Oversees over 20,000 employees in 130 locations worldwide

What we do

- Manage or co-manage six national laboratories for the U.S. Department of Energy.

Brookhaven National Laboratory
Upton, New York

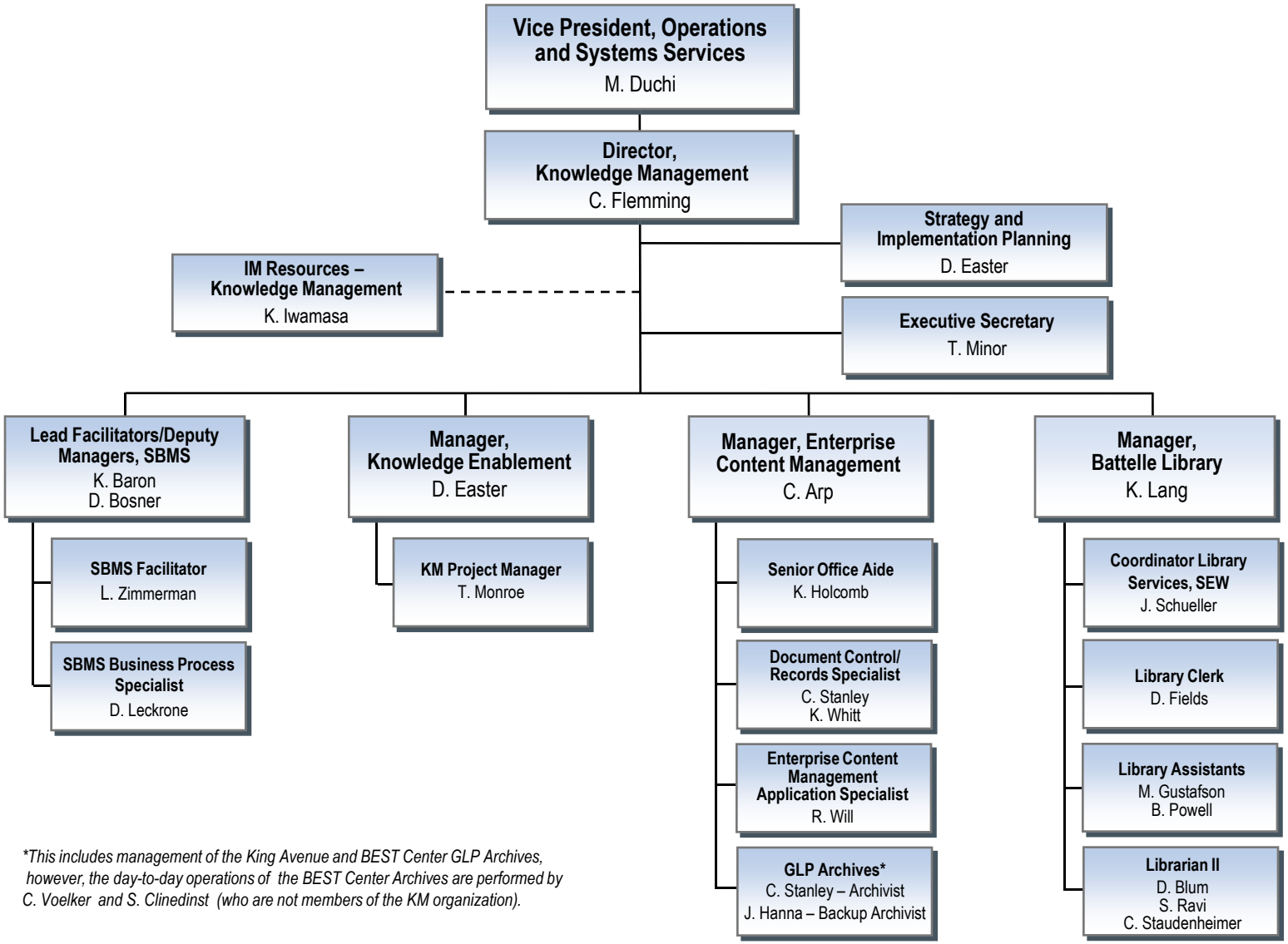
Pacific Northwest National Laboratory
Richland, Washington

Oak Ridge National Laboratory
Oak Ridge, Tennessee

National Renewable Energy Laboratory
Golden, Colorado

Idaho National Laboratory
Idaho Falls, Idaho

Knowledge Management



*This includes management of the King Avenue and BEST Center GLP Archives, however, the day-to-day operations of the BEST Center Archives are performed by C. Voelker and S. Clinedinst (who are not members of the KM organization).

RMO – what we do

- Traditional records management – paper
  - 26,000 cf. in off site storage
- ECM – management of electronic records
  - 8 different groups using our ECM
  - 1.6 million electronic records - 1.4 terabytes
- Purchase, issue and track laboratory records books
- GLP archives
  - Tightly regulated management for FDA and EPA studies
- Vital records program – just getting started

Why harvest SharePoint content?

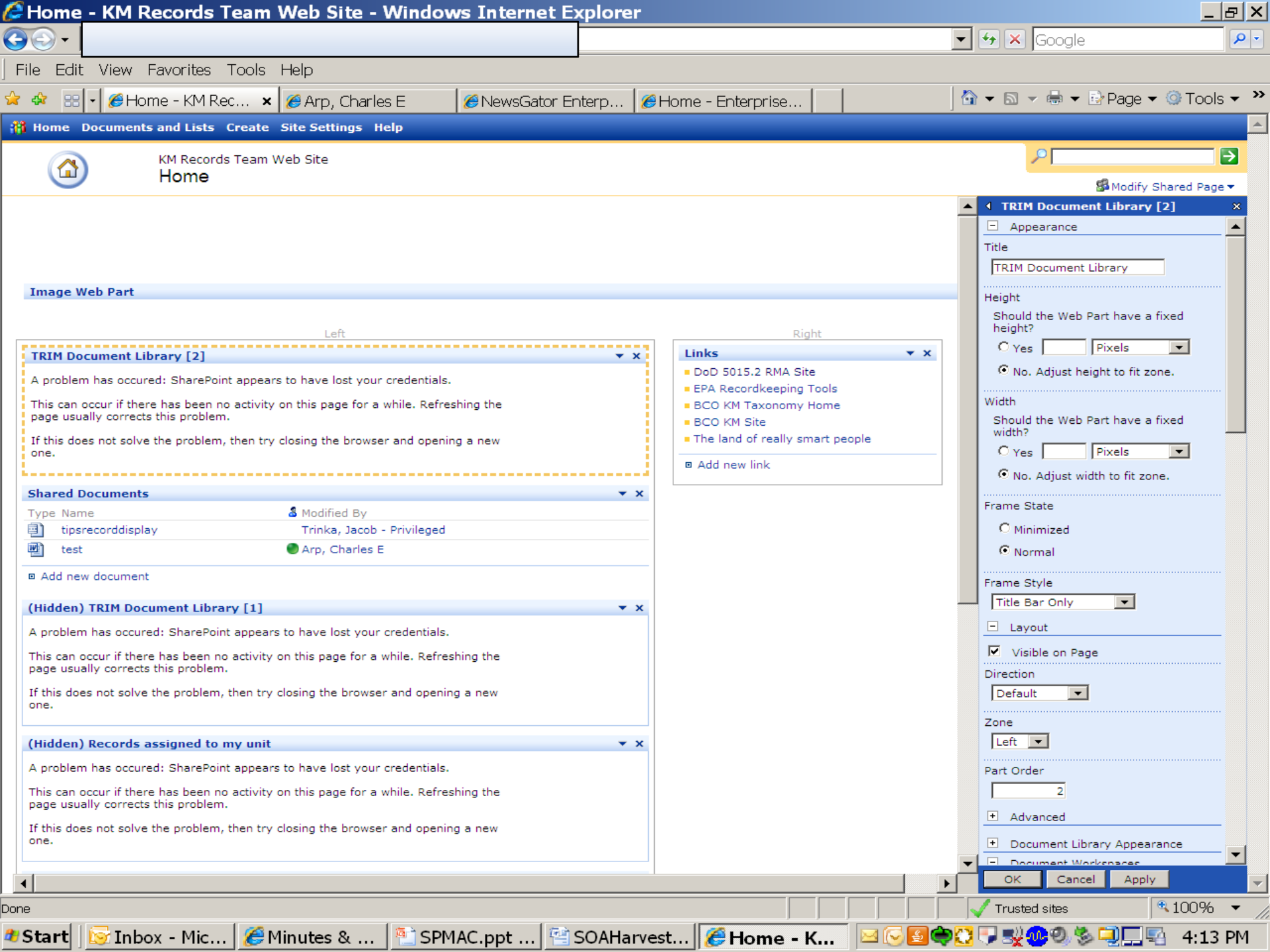
- SharePoint sites - collections of electronic records structured around projects or functions
  - Records managers need to manage the content for legal and/or regulatory purposes
  - Archivists need to appraise and capture the content for historical purposes
- SharePoint content must be captured and ingested into an electronic records keeping repository to manage it over time
  - Records managers use ECM or DoD5015.2 applications
  - Archivists use some sort of digital archive

SharePoint defined

- Microsoft (MS) defines SharePoint as:
  - "an integrated suite of server capabilities…it provides IT… with the platform and tools they need for server administration, application extensibility, and interoperability."
- SharePoint is a collaborative platform
  - It is a web portal that manages lists and digital objects via a database.
  - It uses web parts to modify the content, appearance, and behavior of sites
  - Developers create SP sites with .Net code

SharePoint defined

- Frame off which you hang web-parts
- Three main functionalities
  - Search – ever increasingly powerful search engines
  - Document libraries – documents control, access, versioning
  - List functions – to do lists, discussion lists, calendars, a customizable spreadsheet application
- Strengths are its easy customization, complete acceptance by IT, and propagation of sites


Home - KM Records Team Web Site - Windows Internet Explorer
File Edit View Favorites Tools Help
Home - KM Rec...
Arp, Charles E
NewsGator Enterp...
Home - Enterprise...
Page Tools
Home Documents and Lists Create Site Settings Help
KM Records Team Web Site
Home
Modify Shared Page
Image Web Part
Left
Right
TRIM Document Library [2]
A problem has occured: SharePoint appears to have lost your credentials.
This can occur if there has been no activity on this page for a while. Refreshing the page usually corrects this problem.
If this does not solve the problem, then try closing the browser and opening a new one.
Links
DoD 5015.2 RMA Site
EPA Recordkeeping Tools
BCO KM Taxonomy Home
BCO KM Site
The land of really smart people
Add new link
Shared Documents
Type Name
Modified By
tipsrecorddisplay
Trinka, Jacob - Privileged
test
Arp, Charles E
Add new document
(Hidden) TRIM Document Library [1]
A problem has occured: SharePoint appears to have lost your credentials.
This can occur if there has been no activity on this page for a while. Refreshing the page usually corrects this problem.
If this does not solve the problem, then try closing the browser and opening a new one.
(Hidden) Records assigned to my unit
A problem has occured: SharePoint appears to have lost your credentials.
This can occur if there has been no activity on this page for a while. Refreshing the page usually corrects this problem.
If this does not solve the problem, then try closing the browser and opening a new one.
TRIM Document Library [2]
Appearance
Title
TRIM Document Library
Height
Should the Web Part have a fixed height?
Yes
Pixels
No. Adjust height to fit zone.
Width
Should the Web Part have a fixed width?
Yes
Pixels
No. Adjust width to fit zone.
Frame State
Minimized
Normal
Frame Style
Title Bar Only
Layout
Visible on Page
Direction
Default
Zone
Left
Part Order
2
Advanced
Document Library Appearance
OK
Cancel
Apply
Done
Trusted sites
100%
Start
Inbox - Mic...
Minutes & ...
SPMAC.ppt ...
SOAHarvest...
Home - K...
4:13 PM

Communities and KM

Battelle The Business of Innovation

Communities

All Sites

Communities and KM

This community is the central place for discussions, Q&A, and collaboration on all topics related to Communities and to Knowledge Management within Battelle. The KM Wiki provides information about communities operation and features.

Leave this community
Community security: public
Invite others to join

Tags:
KM Collaboration Content Management Knowledge Sharing Blogging enterprise 2.0 Knowledge Management ECM My Site

Add a tag:

Related Communities

Battelle Women's Network - Community for all Battelle staff members interested in issues and initiatives that impact women.
created February 05, 2009 2:02 PM | 212 members

Young Professionals - Community for young professionals and those interested in generational issues in the workplace
created February 19, 2009 8:55 AM | 196 members

Business Process Management - Community for sharing process

Overview | Discussions | Feeds | Bookmarks | Members | Documents | Calendar | QnA

Recent Community Actions

Tom Monroe added bookmark 'Substituting Social Software Policy for Good Management' to Communities and KM community. — June 10, 2009 9:22 AM

Kathryn Baron added bookmark 'BBC NEWS | Technology | Twitter hype punctured by study' to Communities and KM community. — June 09, 2009 5:45 PM

David Easter added post to thread 'Next Gen Wiki?' in discussion 'General Discussion' to Communities and KM community. — June 04, 2009 3:08 PM

John Kuczek added thread 'Next Gen Wiki?' in discussion 'General Discussion' to Communities and KM community. — June 04, 2009 12:02 PM

KM Community Blog

Toolbar for subscribing to feeds
June 04, 2009 10:55 AM Easter, David A

A Toolbar has been approved for use with Internet Explorer for feed subscription and management. At some point, the Toolbar will be made widely available. However, at present it will have to be subscribed to individually by going to the address and double clicking on the NGToolbarSetup.msi file. This file has been configured for proper security and settings. The Communities wiki contains some additional information on the Toolbar features and functionality.

Actions: Mark Read | Clip Post | Create Discussion | Add Tag |

Working with Bookmarks
May 27, 2009 9:41 AM Monroe, Tom E

Bookmarks can be a very useful feature within communities. Think delicious, only internal and based on communities. There are a few steps you can take, both as a community champion and as a community member, to make using bookmarks easier. You can start by adding a bookmark button to your Internet Explorer toolbar. This allows you to add a bookmark to one of your communities from anywhere on the web. So, if you find a site, article, video clip,

Links

Community Wiki
Wiki Attachments
Community Blog

Add new link

Community Members

See all

Community Tagged News

DynoPlex eOffice Has Google Docs Including Spreadsheet Collaboration – Now $29.95
June 10, 2009 12:00 PM
BerryReview.com [Ronen Halevy]

Yakabod Adds a Social Side to Knowledge Management
June 10, 2009 4:18 AM
Portals and KM [Bill Ives]

InfoSource

Battelle
The Business of Innovation

Announcements

Discussing Smart Grid - Mani Vadari, Battelle's Vice President of Electricity Infrastructure, was one of three panelists March 17 at the Columbus Metropolitan Club's discussion of smart grid. Watch the video of the discussion here.

Kunz Offers Thought Leadership - Download and listen to the podcast from the Research! America 2010 National Forum at the National Press Club in Washington D.C. that HLSGB President Barbara Kunz participated in March 16. She joined other panelists from the FDA, National Institutes of Health and Pfizer, among others to discuss "Translating Discoveries into Better Health."

Battelle Library News - The March edition is now available. Read about the return of *The Chemical Products Synopsis Reports* or check out the Library's new eBooks and eJournal subscriptions. Click here to read the March edition.

Global Climate Change Threat to National Security - Don McConnell, joined former Sen. John Warner (R-VA), Ohio State University professor Andrew Keeler and retired Vice Admiral Dennis McGinn on March 2 to make a presentation at OSU's Faculty Club on how solving global climate change can help ensure America's security. The Pew Project on National Security, Energy and Climate and the John Glenn School of Public Affairs at OSU co-sponsored the forum. View his presentation file here or click here to watch the video of his speech.

Science Fair News - On Thursday, March 11, Battelle's Adrian Felts talked with a local Charlottesville, VA., NBC news affiliate about the importance of STEM education. Manager

Recent Press Releases

Schwope Joins Battelle as Business Development Manager in Armor

Lee Ann Schwope has joined Battelle to help expand its armor and related defense and homeland security business. As a business development manager for Battelle's Industrial & International Market Sector, Schwope will work with the U.S. Department of Defense as well as companies that develop and manufacture vehicles and components.

Read more...

Get to Know

John Cook - holds the keys to all of the activities at Battelle Columbus—literally. As a Lead Security Officer and Locksmith, he oversees every key to every door on the campus. "The keys are kept on hooks associated with room numbers," he says. "I probably have 52,000 keys on hooks down there!" More ...

Flu Information

The *Pandemic Monitor* is a weekly newsletter that provides health updates, including information on the H1N1 virus. Click here to read the most current edition.

Q&A

Click here for information about the H1N1 virus, including details about the recall of the FluMist vaccine.

Flu Vaccination SharePoint Site

My Communities

Battelle Library - Community for Battelle Library users to interact, make suggestions on Library services and materials, and provide feedback to Library staff members.
created September 15, 2009 11:00 AM | 151 members

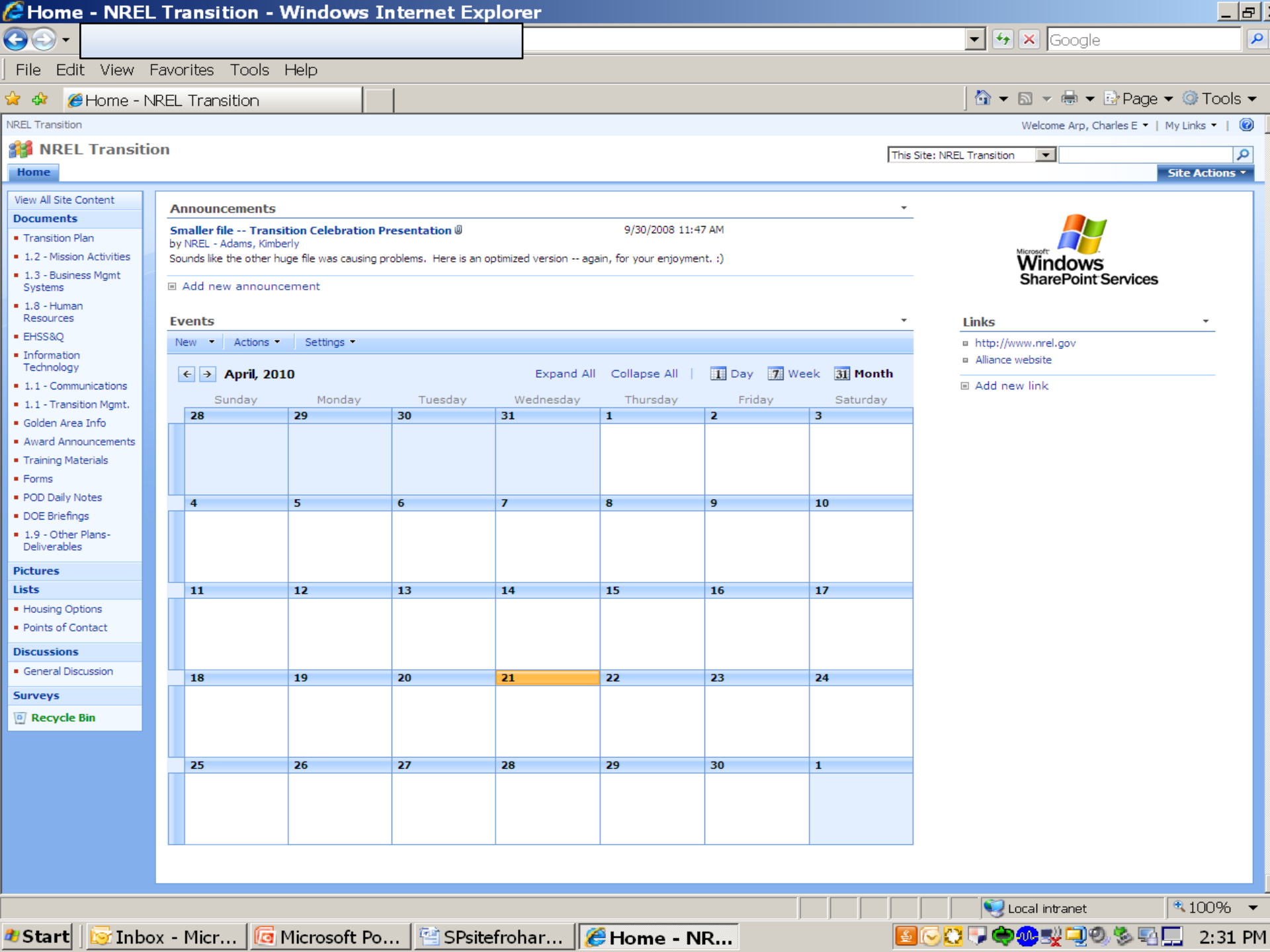
Community Help and Support - Support for Communities, information on enhancements and direction, updates on Battelle's Knowledge Management activities.
created January 26, 2009 12:06 PM | 109 members

KM Group - Private community for Battelle's KM organization
created December 07, 2009 8:56 PM | 22 members

OSS Project Management - OSS Project Management site
created August 11, 2009 4:53 PM | 489 members

SBMS Content Owners - Community of practice for SBMS management system owners, subject matter experts, and points of contact
created June 11, 2009 11:13 PM | 242 members

Unified Communications - Community site for staff members involved in IM's unified communications pilot
created June 15, 2009 9:54 AM | 141 members


Home - NREL Transition - Windows Internet Explorer
Google
File Edit View Favorites Tools Help
Home - NREL Transition
Page Tools
NREL Transition
Welcome Arp, Charles E | My Links |
NREL Transition
This Site: NREL Transition
Home
Site Actions
View All Site Content
Documents
Transition Plan
1.2 - Mission Activities
1.3 - Business Mgmt Systems
1.8 - Human Resources
EHSS&Q
Information Technology
1.1 - Communications
1.1 - Transition Mgmt.
Golden Area Info
Award Announcements
Training Materials
Forms
POD Daily Notes
DOE Briefings
1.9 - Other Plans-Deliverables
Pictures
Lists
Housing Options
Points of Contact
Discussions
General Discussion
Surveys
Recycle Bin
Announcements
Smaller file -- Transition Celebration Presentation
9/30/2008 11:47 AM
by NREL - Adams, Kimberly
Sounds like the other huge file was causing problems. Here is an optimized version -- again, for your enjoyment. :)
Add new announcement
Events
New
Actions
Settings
April, 2010
Expand All
Collapse All
Day
Week
Month
Sunday Monday Tuesday Wednesday Thursday Friday Saturday
28 29 30 31 1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 1
Microsoft Windows SharePoint Services
Links
http://www.nrel.gov
Alliance website
Add new link
Local intranet
100%
Start
Inbox - Micr...
Microsoft Po...
SPsitefrohar...
Home - NR...
2:31 PM

Records Management

Battelle
The Business of Innovation

Navigation

- RMO Home
- Staff Contacts
- Services
- Retention Schedule
- External Resources
- Return To KM

Related Communities

- Communities & KM
- Project Managers

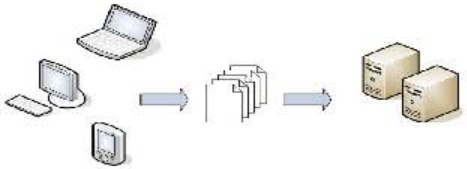
Welcome to the Records Management Site

The Records Management Office (RMO) helps Battelle staff deal with records issues – how to create good records, how long to keep records, and how to store and destroy records. We also issue and manage laboratory record books.

The RMO's mission is to ensure that Battelle records are reliable, complete, cost effective, and accessible for as long as considered necessary by Battelle. The Records Management Office falls under the umbrella of Knowledge Management (KM).

What do you want to do?

- Manage Your Electronic Records in TRIM
- Submit Your Project Deliverable in Electronic Form
- Request a Laboratory Record Book
- Review the Battelle Records Retention Schedule
- Learn more about RMO Services
- Review RMO Procedures in SBMS
- Contact the RMO



RMO Guidelines & Best Practices

The following pages describe best practices and guidelines that you may want to consider in the course of managing records.

- Battelle E-mail Etiquette Guidelines
- Commercial Records Center Selection Guidelines
- Electronic Records
- Guidelines for Managing Email
- Internal Blogging Best Practices
- Inventorying Your Records
- Litigation Hold Procedures

Policy & SBMS Procedures

- RMO Policy 1.3.1 Records Management
- Obtaining, Maintaining, and Transferring a Laboratory Record Book (LRB)
- Retrieving Records From and Returning Records to the RMO
- Transferring Records to the RMO for Storage
- Destroying Records

SBMS Forms & Instructions

- Record of Transfer for Paper Records Form
- Certificate of Records Destruction
- Laboratory Record Book Order Form
- Box Labeling Instructions
- Box Making Instructions
- Box Packing Instructions
- Box Description Example on a Record of Transfer for Paper Records Form

SharePoint History

- 2000 - Windows Server 2000 released
  - "Tahoe" – document management functionality
- 2001 - SharePoint Server 2001(portals) and an add on the Office 2000 called SharePoint Team Services (STS) released
  - STS has web based collaborative functionality (V1)
- 2003 – Windows SharePoint Services (WSS) and Office SharePoint Portal Server 2003 released (V2)
- 2007 - MS Office SharePoint Server 2007 (MOSS2007) released (V3)

SharePoint History

- MS has sold over a million licenses for SP server
- Gartner says “50% of organizations…have piloted or deployed WSS or MOSS 2007 as key elements of their overall information infrastructure.” Gartner, “SharePoint 2010 Steps Up to the ECM Plate”, ID number G00172077
- MOSS 2007 was offered with an add-on records management service pack to meet DoD 5015.2
  - Called “ECM light” in the records management community
  - Limits on number of files in libraries, metadata limitations, no time and event retention triggers, no litigation hold

SharePoint History

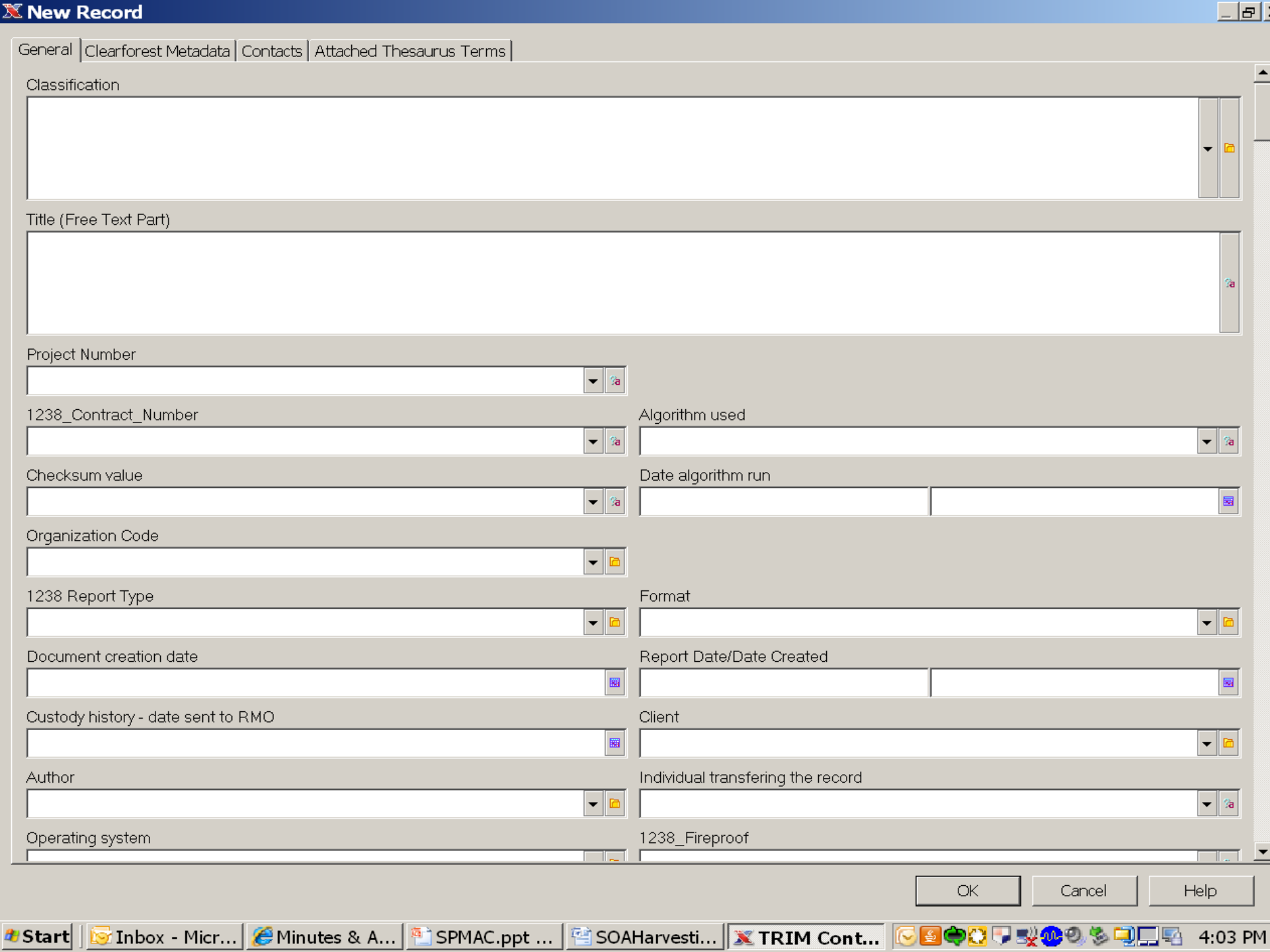
- We had a long discussion over ECM vs. SP
- Our ECM application is HP TRIM
  - Meets all the chapters of DoD5015.2
  - PNNL has used it for over 10 years – over 3 million records in it
- Lots of other good ECM applications out there
  - eDocs (Hummingbird), EMC Documentum, Autonomy Meridio , IBM's Records Manager, Docuware, Open Text's LiveLink , CA's record manager, Iron Mountain's Accutract
- They all have different strengths and weaknesses

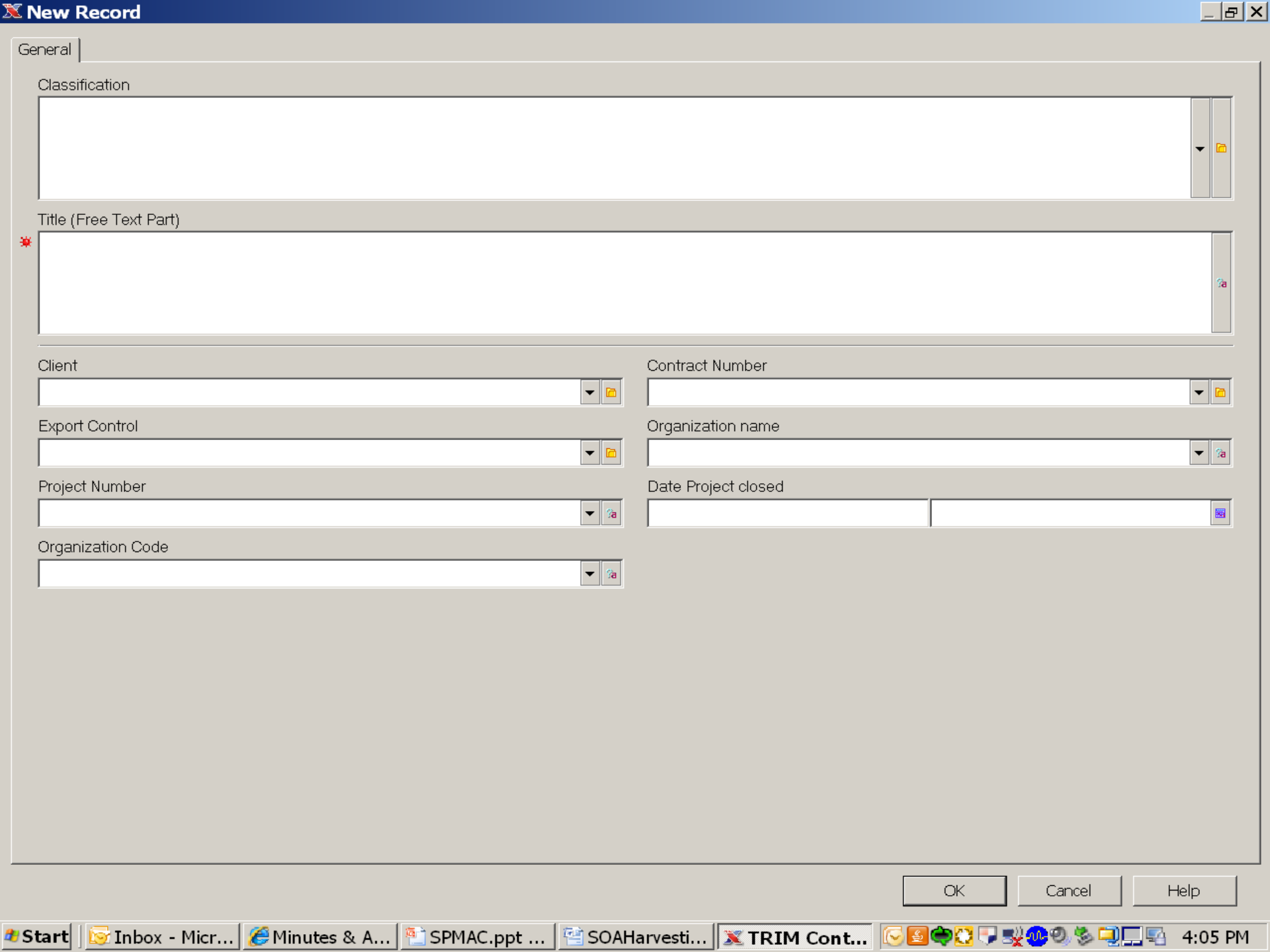
Who is using TRIM?

Commercial	Govt	International
Honeywell	Army, Air Force	UN
Merck	Navy/Marine Corps	Govt of Brit Columbia
AstraZeneca	DARPA	Dutch National Bank
King Pharma	NRO	UK Dept. Trade & Industry
Mercedes Benz	NARA	Most of the Australian Govt.
AT&T	FDIC	South African Defense
National Geographic	Office of SecDef	Ireland Civil Ser.(18,000)
NASCAR	EUCOM	30 different countries
MGM	FBI, IRS, FEMA	
Penn Power & Light	GA State Archives	
Charles Schwab	CENTCOM	
Papa Johns	DOE – PNNL, SNL, Yucca Mt.	

Harvesting Content

- We harvest three records types
  - Final reports, project management records, and working files
- Final reports are harvested as individual files
  - Specific record type in TRIM, adding metadata to PDFs
- Project management and working files – we create a folder in TRIM and apply metadata to the folder
  - All the files within the folder inherit the metadata
- Use a workflow to connect a network folder to a specific project folder in TRIM
  - Dump the captured SP files into the network folder


New Record
General
Clearforest Metadata
Contacts
Attached Thesaurus Terms
Classification
Title (Free Text Part)
Project Number
1238_Contract_Number
Algorithm used
Checksum value
Date algorithm run
Organization Code
1238 Report Type
Format
Document creation date
Report Date/Date Created
Custody history - date sent to RMO
Client
Author
Individual transfering the record
Operating system
1238_Fireproof
OK
Cancel
Help
Start
Inbox - Micr...
Minutes & A...
SPMAC.ppt ...
SOAHarvesti...
TRIM Cont...
4:03 PM


New Record
General
Classification
Title (Free Text Part)
Client
Contract Number
Export Control
Organization name
Project Number
Date Project closed
Organization Code
OK
Cancel
Help
Start
Inbox - Micr...
Minutes & A...
SPMAC.ppt ...
SOAHarvesti...
TRIM Cont...
4:05 PM

Harvesting Content

- Which sites get captured?
  - Site administrator contacts the SharePoint team and says the site needs to be “archived”
  - Orphaned sites are identified by the SharePoint team and sent to the RMO on a annual basis
- Site administrators copy the site user/permission list as a Word or Excel document and places it in a library
- Any list that the site administrator wants captured is copied into a Word or Excel document and placed in a library
  - Cannot capture lists or list attachments unless they have been copied to Word and placed in a library

Harvesting Content

- IM sends the RMO a link to the site that is to be archived
  - They lock down the site prior to sending us the link
  - They make RMO staff the site administrator
- For each document library within the site:
  - Hit the “Shared Documents” link then the “explorer view”
  - Select and copy the files to the network folder linked to TRIM
  - The files are automatically placed in the designated TRIM project folder
- RMO deletes the site after we have copied the libraries

Harvesting Content

- Problems:
  - Too many sites to be harvested in this fashion
    - Over 6,000 SP sites at Battelle
    - Too much work for the administrators and for the RMO
    - SharePoint team needs to identify and send us orphaned sites
  - Must have the ability to harvest list items - blogs and wikis
    - List item attachments
  - Too easy to miss content within a site
    - Site designs - Rubik's cube
    - Customized labels
  - Not getting the look and feel of the site
    - Is it needed?

Harvesting Content

- Problems
  - Labels on document libraries
  - Folders within folders within folders…
  - Reusing file names within different folders
    - Monthly report in May folder, Monthly report in June folder…
  - Lack of detail on the sites – “this is our site…”
  - Size limitation on copying files – 56,000 KB
- Solutions
  - Input into the SharePoint “Governance document”
    - The rules for implementing SharePoint at your institution
    - Written by the IT SharePoint team

What's next – SharePoint 2010

- In beta release now – has an ECM module
  - Release scheduled for this fall
- MS says it has tools for:
  - Retention
  - Legal holds
  - Metadata – easy to add and has tools for automatic capture
  - All content published via the SharePoint server can be managed
    - Includes blogs, wikis, comments, ratings, user defined taxonomies
- No mention of meeting DoD5015.2

What's next – SharePoint 2010

- "Organizations looking to SharePoint to support large volumes of static content or transactional processes will find their needs better met through partner-built solutions extending SharePoint or competing ECM offerings" Gartner, "SharePoint 2010 Steps Up to the ECM Plate", ID number G00172077

- Gartner and others have reported that they believe that SharePoint 2010 will not attempt to meet DoD5015.2 record keeping standards

- Our plan has always been to use SharePoint, TRIM, and Clear Forest in harmony

TRIM 7

- 64 bit application in Unicode
- Better search
- Web client has zero footprint
- Can archive all SharePoint content – including list items with attachments (blogs, wikis, to do lists)
- Has an “archive” module that lets you de-hydrate and re-hydrate sites
  - Automated process kicked off with pre-defined rules
  - Re-hydrating a site put content on a SP template
  - Look and feel?

TRIM 7

- We want/need to control records during their active life – we need TRIM/SharePoint integration module
- TRIM/SharePoint integration uses the SharePoint interface to place records in TRIM, and to search and retrieve records from TRIM
- Four different integration modes with TRIM/SP
  - Managed – file in TRIM & SP, can edit file/metadata in SP
  - Finalized – file in TRIM, metadata in SP, file and metadata cannot be edited
  - Relocated – file in TRIM, no metadata in SP, no edits
  - Archived – file in TRIM – no edits, no access from SP

ECM applications and SharePoint

- One of the biggest issues for us is going to be configuring the template for generating TRIM/SharePoint sites
  - And the SharePoint governance document
- The TRIM 7 SharePoint integration module is impressive – but lots of other ECM applications have integration with SharePoint
- AIIM has a nice site devoted to ECM applications that connect with and enhance SharePoint. http://www.aiim.org/sharepoint/

Conclusions

- Archivists and Records Managers must be able to harvest SharePoint sites
  - Their deployment is too wide spread to ignore
- Because of the popularity of SharePoint, ECM tools and integration modules are being created
  - Metadata creation and editing is common to all
- SharePoint archiving tools are evolving
  - Automated capture is coming
  - The ability to capture look and feel still evades us
- Jump in the pool and play

Questions?

arpc@battelle.org
(614) 424-7897

Thanks for your time