

Automated Records Management



Transformation
through Partnerships

DocMan Electronic Content Disposition System (eCDS)

Steve VonVital, DOE EERE CIO

April 16, 2012

Electronic Record Challenges

- Many agencies have ineffective “print and file” policies in place that do not account for different types of media.
- Email poses significant challenges to meeting federal record keeping obligations.



Electronic Record Challenges

- Records are typically stored on email servers, public drives and on the end user's local drives. The result is replicated and redundant files stored across multiple locations.
- Without the ability to organize and categorize this massive amount of information, IT departments cannot delete any electronic files for fear of legal or regulatory repercussions.

By Alice Lipowicz

Thursday, March 3, 2011

Agencies admit to bad records management

“Many agencies do not ensure that e-mails are preserved consistently and do not monitor compliance. Furthermore, many agencies use inappropriate preservation strategies, such as using system backup as a means of preservation, or printing e-mails on paper.”

many
electronic records, and
resources dedicated for that purpose,
NARA said.

preservation strategies, such as using system
backup as a means of preservation, or printing e-
mails on paper.

1. Content Management

2. Records Repository





3. Automatic Categorization




Watch the Process


Watch the Process





Site Actions  Browse Page

Keating-aitken, Ann (CONTR) ▾

 Records Center I Like It Tags & Notes

Records Center EERE Records - FILES EERE Record Center Search this site...  


Libraries

Drop Off Library

Record Library


Recycle Bin


All Site Content



Records Center

Use this site to submit and find important documents that should be stored for long term archival. Submitted records are automatically placed in the correct library and folder based upon the properties you fill out.

 Submit a Record

UKWHND2VNH3H-1-1 
Find by Document ID

Use this web part to teach visitors to the Records Center about records management and organizational compliance.


Add information about records management topics, such as:

- The definition of a record in your organization
- What happens to a record after it is submitted to the Records Center
- Steps users can take to comply with organizational policy

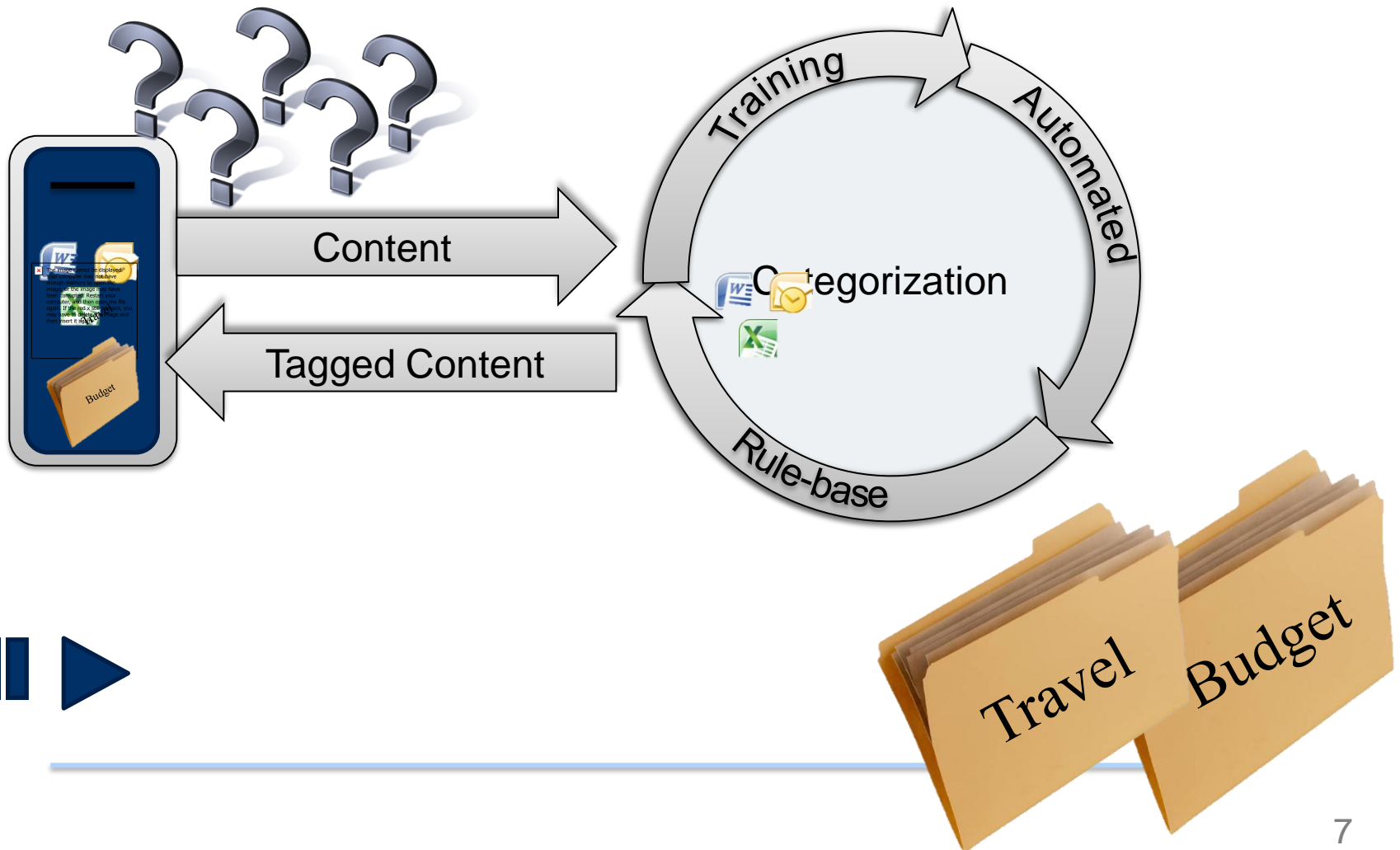
Add links to other organizational compliance sites, such as:

- Your organization's compliance training site
- A site about organizational retention policies
- A list of records management contacts for each department

My Records Pending Submission

Type	Name	<input type="checkbox"/> Modified By
There are no items to show in this view of the "Drop Off Library" document library. To add a new item, click "Add document".		
 Add document		

Automatic Categorization



- SharePoint 2010
 - ✓ Email – Utilize Exchange 2010 Journaling
 - ✓ SharePoint Sites and Content
 - ✓ Shared Drives (*Future*)
 - ✓ Local User Drives – Migrate content to external storage & SharePoint My Sites (*Future*)

Component #2

Electronic Records Repository

- Electronic emails and files are automatically routed to the SharePoint Records Center and categorized by Recommind.
- Simple for Records Managers to learn and use.
- Record Managers can configure multi-phase disposition schedules.

Component #3

Automatic Categorization

- Allows thousands of newly-created electronic records to be classified daily.
- Ensures that email-based information is properly tagged and categorized with **no impact** on busy professionals.





- Machine learning offers the highest potential for automatic categorization accuracy (as more records are added the accuracy increases).
- Recommind uses a patented algorithm known as Probabilistic Latent Semantic Analysis (PLSA).
- Decisiv identifies and structures relevant concepts and topics within record training sets.

- **PHASE I**

- Record Collection

- **PHASE II**

- Record Training

- **PHASE III**

- Ongoing Training and Testing

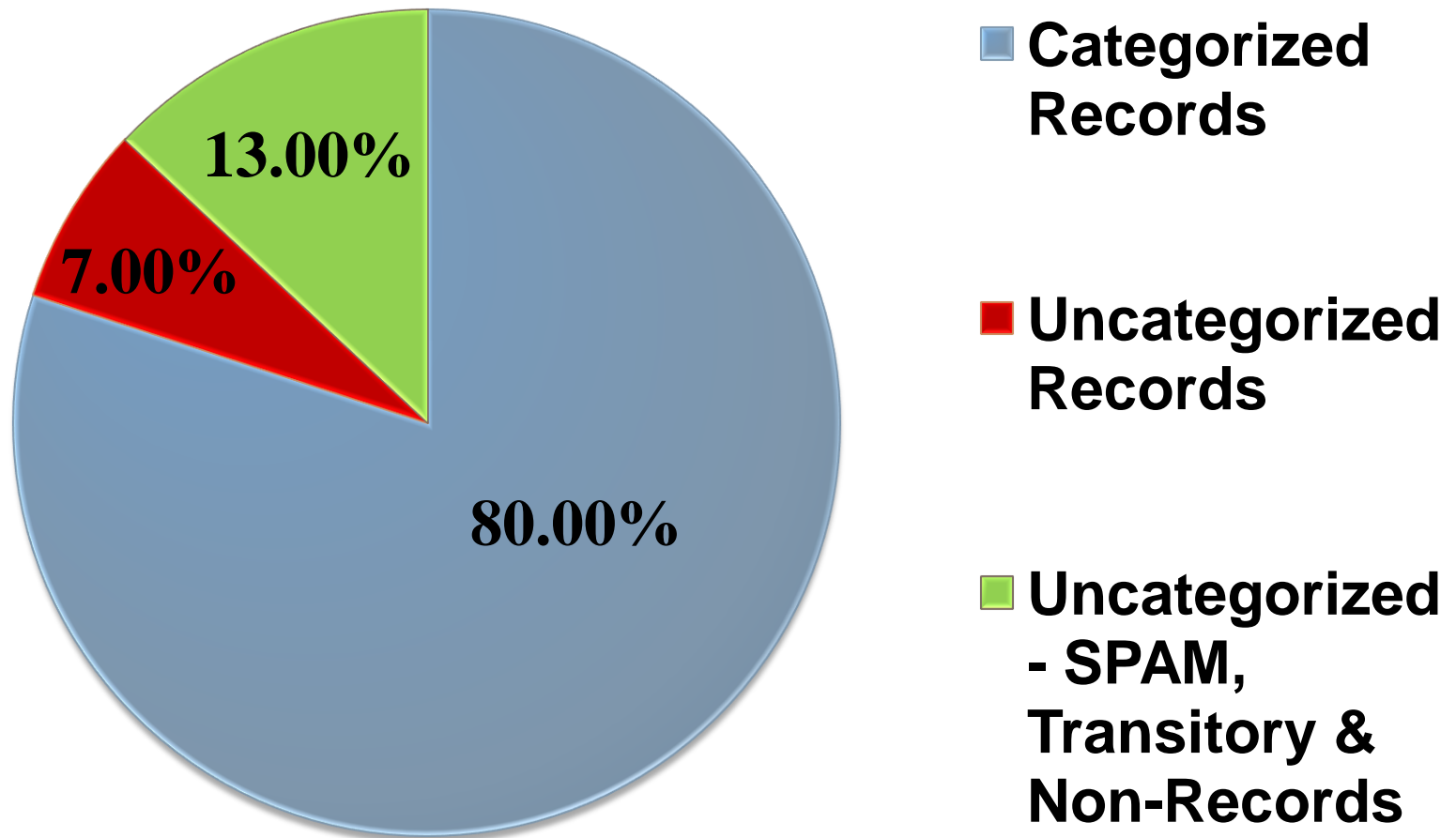
- **CATEGORIZATION RATE**

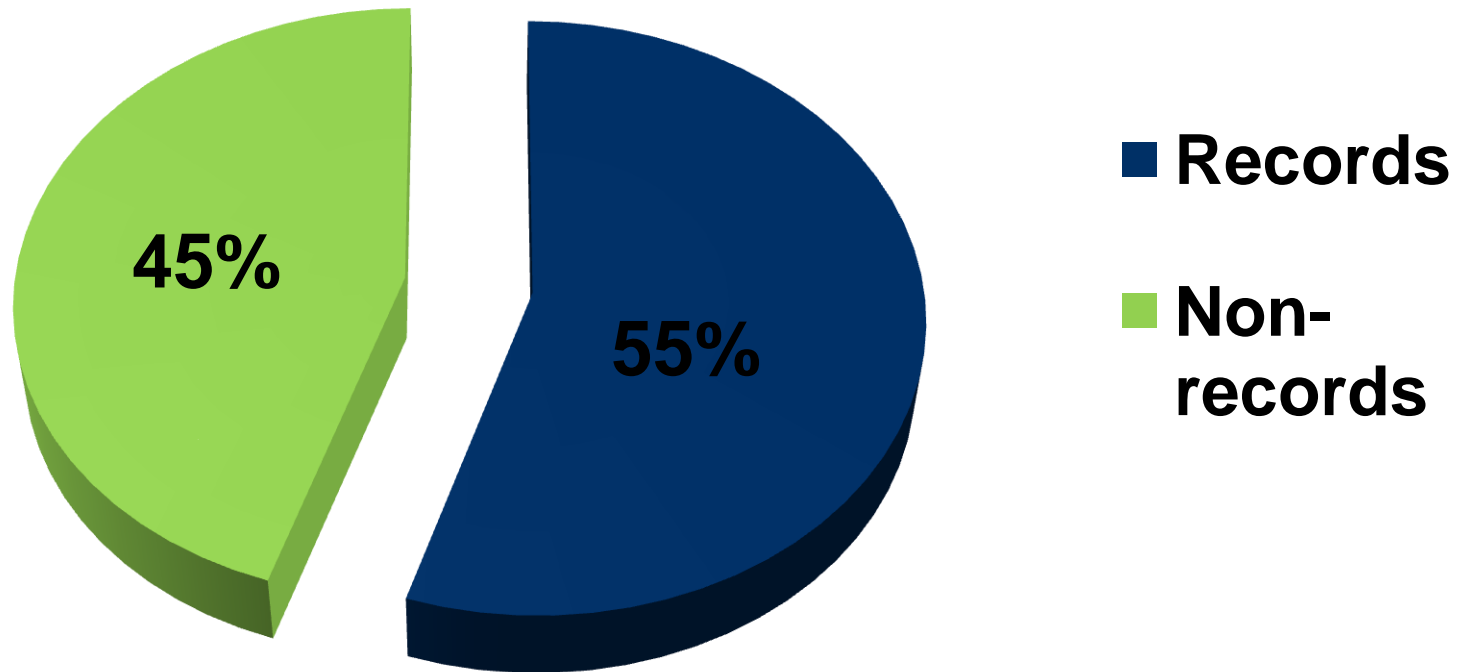
- Percentage of total files categorized.
 - *Monitor production system.*

- **CATEGORIZATION ACCURACY**

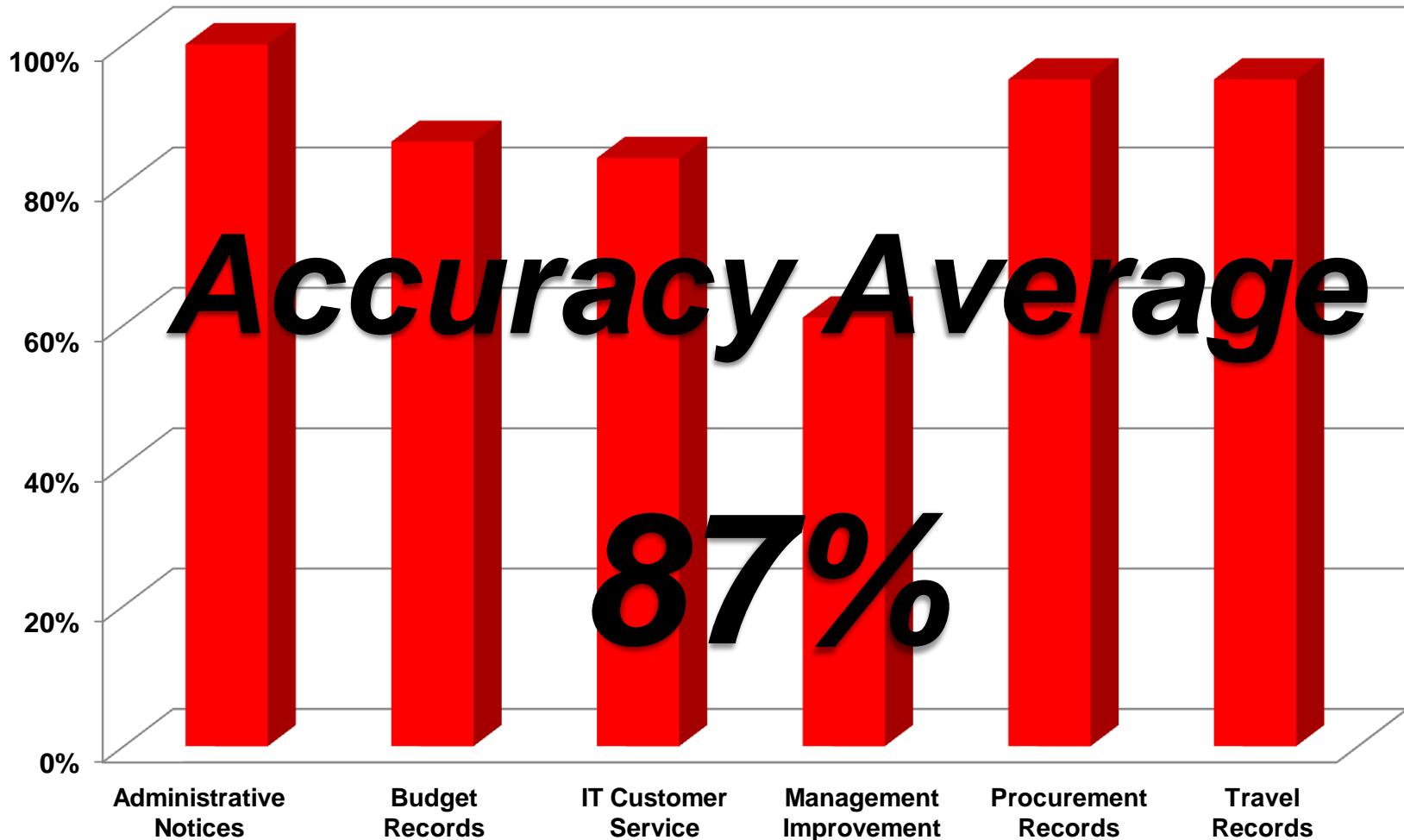
- Precision
- Recall
 - *Controlled test groups on sandbox.*

Email Categorization Rate





Email Categorization Accuracy



- **Over 1 million files (1.6 TB).**
- **30% categorized and audited during Phase II.**
- **Average accuracy is 82%.**
- **Legacy data is the biggest challenge.**

The New York Times

By John Markoff

Friday, March 4, 2011

Armies of Expensive Lawyers, Replaced by Cheaper Software

“The computers seem to be good at their new jobs. Mr. Herr, the former chemical company lawyer, used e-discovery software to reanalyze work his company’s lawyers did in the 1980s and ’90s. His human colleagues had been only 60 percent accurate, he found.”

Some programs go beyond finding documents with relevant terms at computer speeds.

powerful software programs replace work once done by legions of logic designers and draftsmen.

“There is no reason to think that technology creates unemployment,” Professor Author said.

- 800+ users at headquarters are submitting email through the system.
- Over 30,000 email messages are categorized daily.
- Shared drive categorization is currently in Phase II.

- Migrate local drives to SAN for categorization.
- Convert paper records to electronic files for categorization.
- Use Recommind Axcelerate for FOIAs and e-Discovery.
- Implement additional Recommind modules:
 - File collection from multiple data sources.

- **EMAIL**

- Official records are maintained in the SharePoint Records Center.
- Email stored on Exchange servers will be deleted after one year.
- Legacy email to be collected and categorized in the Records Center.

- **SHARED DRIVES AND SHAREPOINT SITES**

- Records on the shared drives are crawled and categorized.
- SharePoint content will be locked down in place after one year of inactivity. Records will be transferred to Record Center and categorized at appropriate time.

U.S. Department of Energy

(Office of Energy Efficiency and Renewable Energy)

Steve VonVital, CIO
steven.vonvital@ee.doe.gov
202-586-2978