Using TAR To Improve RIM
David White & Matt Cohen – AlixPartners LLP.
Feb 13th, 2013
Introduction

David J. White
Director, AlixPartners LLC

David is a Director in the eDiscovery and Information Governance practice group at AlixPartners. His practice focuses on issues regarding Information Lifecycle Governance with a particular focus on electronic discovery and international and domestic data privacy and security. He has nearly two decades of experience assisting corporations in multiple sectors in preparing for and responding to discovery in litigation and regulatory matters, including electronic document preservation, production, spoliation mitigation, and computer forensic investigations. Prior to joining AlixPartners, he was a Partner in the Commercial Litigation practice group at a top AmLaw100 law firm. David is a certified Six Sigma Green Belt, and Certified Information Privacy Professional (CIPP/E/US), and a registered US Patent Attorney. He also holds a second law degree from London University (LSE).
Can TAR Be Used To Improve RIM?

RIM & ESI

- The Rapid increase in data volumes is creating greater disconnects between ESI and Records Retention Schedules

TAR

- What Is It? How it works?

A Different Process

- How TAR can be used to locate and classify records in large ESI repositories
RIM and ESI
We have Reached a Tipping Point

The growth rate of information now far exceeds IT budgets and the processes for governing that information.

IT departments everywhere are shedding costs as rapidly as possible, and reducing IT spend as a percentage of revenue to align with revenue compression.

Accumulating, storing and managing data without value is no longer rational for the company or shareholders.

Companies are working to define and operationalize programs and improve processes that enable defensible disposal of unnecessary data.

This effort can curb storage growth, dramatically lower IT RIM and legal costs and lower the organization’s risk profile systemically.

90% of the data in the world was created in the last two years.*
Excess Data Creates Excess Cost & Risk

Data with no business value or legal duty

Avg 50% growth every year

How does RIM define their piece of the pie?
The ESI Storage & RIM Disconnect: Where Have All The File Clerks Gone?

**Ideal**
Classify and code records at the time of creation and manage accordingly

**Actual**
Create, store, retain, with little or no organization and try and figure it out later.

77% of companies report that current schedules aren’t actionable on electronic information in their current form.
Meanwhile IT Is Taking Action

Take out all unnecessary costs…

- Decommission as many apps and servers as fast as possible
- Disposing of unnecessary data, storage and back up tapes
- Outsource storage to cloud service providers…
- But no worries about legal … because they only care about email and hard drives
- Full Steam Ahead without considering Records Retention obligations or policies (“what retention policy?” “don’t those only apply to paper?”)
- Mergers, divestitures & RIFs only compound problem
Technology Assisted Review (TAR): What Is It? How Is It Currently Used?
The Evolution To Technology Assisted Review in Discovery

Paper to Digital

Linear Manual Review

Boolean “KWS” circa 2007

Clusters & Conceptual Search

Predictive Coding

TAR

Massive Cost & Time. Plus Human Error

Higher Levels of Accuracy and Lower Costs

Estimates at only 65% accuracy for both Recall & Precision
Technology Assisted Review ("TAR") – What Is It?

What It Is:

It is an analytical approach to a population of data that uses complex algorithms to reliably propagate decisions about small sample sets of documents across larger populations. A workflow that uses:

- Conceptual Search Algorithms
- Content Clustering
- Statistical Analysis
- Predictive Coding

What It Isn’t:

- Black Box Voodoo
- Blind Reliance on Technology
- A Substitution For Critical Human Analysis
How TAR Can Be Used To Locate & Classify Records In Large ESI Repositories
Use of TAR in RIM applications

- Categorization: Data that is currently in unstructured systems
- Classification: More accurate adherence to GARP
- Data Disposition: Dispose of Date At Expiration
Two Approaches to Classification

Metadata Analysis
- File Types
- MAC Dates
- Creators/Users
- File Paths

Content Analytics
- Content Clustering
- Predictive Coding
Information Value Declines Over Time While Risk and Cost Increase
A Sample Process

Utilize a seed set of information to determine how the company’s actual information record types compares with the company’s records taxonomy.

Use of a TAR process across company ESI repositories will enable a predictive capability for the classification of documents helping records to more accurately make decisions on classification policy.

Once the new system for classification is defined the process can be aimed at other storage locations identified by a data map as identified on the next slide to dramatic results.
Data Map By Data Source/Location/Type – Applying TAR techniques

Email

Documents

System Files/Noise

Reports-Plans etc.

Sample ESI Documents using predictive techniques. Compare/Contrast with Records Retention Schedules and legal holds.

Refine retention and archive strategies accordingly, and defensibly delete waste.

Adhere to Obligations Matrix
ESI Analysis – Volume & Location
ESI Analysis – Last Access Date

Select Infoset
- All Objects Infoset

Group By
- None
- By Source Type

View By
- Last Accessed View
  - 1 to 6 months
  - 6 to 12 months
  - 1 to 3 years
  - 3 to 5 years
  - 5 to 7 years
  - Over 7 years

Select Data Overlay
- No Overlay
  - Select an Overlay to reveal information about your data.

Data Map
- Last Update: 4/26/22 13:56
- Object Count: Size On Disk

All Infoset
- Exchange
  - 1-6 months
  - 1-3 years
  - 3-5 years
  - 5-7 y.
  - 6-12 months
  - 6-12 m.

SharePoint
- 1-6 months
- 1-3 years

Centera
- 1-6 months
- 1-3 years

CIFS
- 1-6 months
- 1-3 years

FileNet
- 1-6 months
- 1-3 years

NFS
- 6-12 m.
- 6-12 mon.
- Over 7

Data Map Details
- Over 7 Years, Top Data Sources
  - Exchange Austin 1
  - 2634
  - Data Source 1
  - 3401
  - Data Source 45
  - 3278
  - Data Source 33
  - 3331
  - Data Source 4
  - 1833
  - Data Source 51
  - 1852

Export Object List

Did you know?
The View By menu lets you set custom views on your infosets.
ESI Analysis – PII/PCI Compliance
Taking Action – Clean Up

Selected Infoset: All European Email
- Number of Data Objects: 124,563,299
- Size on Disk: 475,236 MB
- Created: 4/12/12 3:45pm
- Updated: N/A
- Description: All email collected from European Exchange data sources.

Available Actions:
Select an Action from the Actions below: Learn More...

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email actions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archive all email</td>
<td>Copy to Retention</td>
<td>Copies all email over 90 days old to</td>
</tr>
<tr>
<td>Delete personal email</td>
<td>Delete</td>
<td>Deletes all email sent from non-com</td>
</tr>
<tr>
<td>File actions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete aged files</td>
<td>Delete</td>
<td>Finds and deletes all files older than</td>
</tr>
<tr>
<td>Move PII</td>
<td>Move to Retention</td>
<td>Moves files with PII to secure location</td>
</tr>
<tr>
<td>Legal Collection</td>
<td>Copy to Litigation Hold</td>
<td>Collect files to legal retention repository</td>
</tr>
</tbody>
</table>

Action Details:
- Name: Move PII
- Description: Moves files with PII to secure location
- Type: Move to Retention
- Last Used On: Infoset1, Infoset2, Infoset3, Infoset4

Run Action
Closing the Disconnect by Data Mapping

Conduct data inventories to identify information organizations have, where they keep it, and the retention requirements.

Define data-source specific retention policies tying obligations and value to assets.

Enable local schedule variations as needed; compare retention practices for the same types of information across business, systems, and repositories.
Sample catalogue of data source with legal holds, retention schedules, privacy and confidentiality requirements.
TAR Utility Summary

- Determining what you have and where it is.
- Cleaning-up debris and de-duplicating data.
- Analyzing content and assessing patterns to define and apply appropriate policies.
- Migrating legacy data to the appropriate storage tier for greater cost efficiency.
- Auditing and enforcing retention policies, and dispositioning data beyond mandated retention schedules.
Questions?
Global Locations

AlixPartners is ready to field a team of relevant experts whenever and wherever they are needed. Our professionals work from 16 global offices in more than a dozen different countries. They speak more than 50 languages, and have experience in every corner of the world. Call us, we'll be there when it really matters.