Privacy Through Transparency

Brian Tuemmler
Nuix
Agenda

• Information Governance
• Index and classification lakes and maps
• Content cleanup
• Data privacy mandates
• Breach notifications and regulatory response
What is the paradox?

• Why do criminals seem to be able to see your exploitable content better than you
Governance Conflicts

- Versioning vs Concurrency
- Preservation vs Retention
- Age vs Relevance
- Security vs Administration
- Sharing vs Protecting

Source: IGRM www.edrm.net
Common IG Initiatives

- Global compliance framework
- Litigation ready lake
- Digital transformation
- Federated records management
- Enterprise Information Management
- Consolidated IG platform
Policy Framework

NUIX IG EXECUTION FLOW

POLICY

TECHNOLOGY

TRAINING AND COMMS

PROCESS

AUDIT

METRICS

CONTINUOUS IMPROVEMENT

TRADITIONAL ROADBLOCK
Common use of Evidence versus Testimony

• **Evidence**
  - ✓ Litigation
  - ✓ Investigation

Proactive use of evidence is underutilized for many common business functions – especially those fostering ethical behavior

• **Testimony**
  - ✓ Litigation
  - ✓ Investigation
  - ✓ Records Management
  - ✓ Cyber security
  - ✓ Mergers and Divestitures
  - ✓ Risk Management
  - ✓ Business Change
Index and Classification Software

Index  Understand  Classify  Act  Monitor
Indexed Data Lake

• **Fully searchable repository** of your entire enterprise-wide user generated content. Leveraging large data sets for BI, policy management, IT optimization, business driven change
  
  • Proactive business intelligence
  • Address multiple mandates and unique use cases
  • Use the evidence in your content for internal information transparency

Finding the single source of truth through total data intelligence
Map Content

• Shows where you are and where you need to go
• Inventory and high level classification of corporate assets
• Iterative process as required
  • No text, few settings for speed
  • Incrementally add for depth
  • Targeted update of index to meet risk criteria
Privacy Data Map

• Identify where to find and not find data
  • By technology/platform location
  • By geospatial locations
  • By type of data
  • By level of access
  • By function

• Leverage sources, file types, metadata, subject matter, number patterns
Typical Enterprise Unstructured Content

- Duplicates and Trivial: 20%
- Databases: 7%
- Applications: 12%
- Web Content: 5%
- Multimedia: 14%
- Templates: 2%
- Other: 40%

Challenges - 50%:
- Audio files
- Video files
- Image files
- Databases
- Cloud sources
- Uncommon formats
- Emails

Classified by semantic analysis
Identify

Finance Marketing HR IT

NAS02006 NAS01772 NAS11266
Classify

• Classify content in the indexed data lake
  • Meet retention, value, risk requirements
  • Respond efficiently
  • Holistic – classify what is, and is not

• Keywords, patterns, topics, similarity, predictive coding, context
For RIM

• An indexed data lake can give you:
  • What category provides greatest storage impact benefits
  • What categories apply to what sites and shares
  • LOE and benefit from purging
  • Additional clarification and exemplars for categories
  • Who owns the content to identify custodians
Prioritize classification efforts

- Not all classifications will contain equal volumes or importance
- 10 categories may represent 80% of expired records

### Row Labels

<table>
<thead>
<tr>
<th>Event</th>
<th>Retention</th>
<th>Expired</th>
<th>2,012</th>
<th>2,011</th>
<th>2,010</th>
<th>2,009</th>
<th>2,008</th>
<th>2,007</th>
<th>2,006</th>
<th>2,005</th>
<th>2,004</th>
<th>2,003</th>
<th>2,002</th>
<th>2,001</th>
<th>2,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNV2020 - Retain drafts and working files no longer than completion of the final version</td>
<td>2</td>
<td>-</td>
<td>17,702</td>
<td>5,343</td>
<td>9,483</td>
<td>7,362</td>
<td>4,771</td>
<td>2,180</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHN3040 - 7 Years</td>
<td>7</td>
<td>9,792</td>
<td>5,789</td>
<td>13,507</td>
<td>9,393</td>
<td>8,949</td>
<td>4,499</td>
<td>1,51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRC4000 - No Longer than 3 Years</td>
<td>3</td>
<td>7,986</td>
<td>1,462</td>
<td>2,119</td>
<td>1,587</td>
<td>908</td>
<td>389</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP9900 - No Longer than 5 years; recommended retention of up to 3 years</td>
<td>-</td>
<td>5</td>
<td>6,101</td>
<td>114</td>
<td>577</td>
<td>650</td>
<td>695</td>
<td>701</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCR2010 - Contract Execution + 10 Years</td>
<td>5</td>
<td>10</td>
<td>3,836</td>
<td>4,970</td>
<td>3,947</td>
<td>4,317</td>
<td>970</td>
<td>1,346</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMK1000 - Last Authorized Use + 6 Years</td>
<td>3</td>
<td>6</td>
<td>3,325</td>
<td>7,922</td>
<td>12,515</td>
<td>4,418</td>
<td>2,281</td>
<td>1,588</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP8000 - No Longer than 5 Years</td>
<td>-</td>
<td>5</td>
<td>2,288</td>
<td>315</td>
<td>491</td>
<td>616</td>
<td>1,248</td>
<td>574</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCR7000 - Expiration + 6 Years</td>
<td>5</td>
<td>6</td>
<td>2,140</td>
<td>255</td>
<td>142</td>
<td>34</td>
<td>254</td>
<td>295</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCR3200 - Resolution + 10 Years</td>
<td>2</td>
<td>10</td>
<td>1,906</td>
<td>3,947</td>
<td>4,317</td>
<td>9,496</td>
<td>7429</td>
<td>9,795</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPS2000 - 10 Years</td>
<td>10</td>
<td>1,598</td>
<td>6,904</td>
<td>8,746</td>
<td>8,938</td>
<td>7429</td>
<td>9,795</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITS1300 - Until system is no longer relevant to existing data</td>
<td>3</td>
<td>10</td>
<td>1,248</td>
<td>773</td>
<td>587</td>
<td>197</td>
<td>389</td>
<td>421</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPS1300 - 10 Years</td>
<td>10</td>
<td>1,111</td>
<td>592</td>
<td>727</td>
<td>1,061</td>
<td>863</td>
<td>408</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPI5300 - Date of Use + 6 years</td>
<td>1</td>
<td>6</td>
<td>992</td>
<td>1,186</td>
<td>1,444</td>
<td>1,165</td>
<td>2,722</td>
<td>3,045</td>
<td>2,62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP1000 - 10 Years</td>
<td>-</td>
<td>10</td>
<td>988</td>
<td>2,657</td>
<td>3,439</td>
<td>1,650</td>
<td>3,417</td>
<td>21,644</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITS3210 - 90 days to 3 Years</td>
<td>3</td>
<td>957</td>
<td>1,411</td>
<td>914</td>
<td>899</td>
<td>548</td>
<td>301</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP7000 - 10 Years</td>
<td>-</td>
<td>10</td>
<td>714</td>
<td>453</td>
<td>1,208</td>
<td>1,197</td>
<td>137</td>
<td>405</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPS1200 - 6 Years</td>
<td>6</td>
<td>701</td>
<td>3,018</td>
<td>2,092</td>
<td>1,004</td>
<td>1,156</td>
<td>1,430</td>
<td>2,42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPI2000 - High priority</td>
<td>3</td>
<td>6</td>
<td>682</td>
<td>283</td>
<td>491</td>
<td>333</td>
<td>936</td>
<td>416</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

80% compliance
Content Cleanup

1. Govern your data
2. Determine rules and roles
3. Preserve
4. Put away
5. Cleanup
   - Unilateral – Enterprise wide consistent policy based decisions
   - Uniform – RM Categories, workgroup
   - Unified – Review by the custodian or owner
6. Address larger IG issues
Current practices

What do people use shared drives for?

<table>
<thead>
<tr>
<th>Works well in ECM Program</th>
<th>Doesn’t work well in ECM</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Capture important data</td>
<td>• Backup files or folders</td>
</tr>
<tr>
<td>• Create compound documents</td>
<td>• Develop software code</td>
</tr>
<tr>
<td>• Collaborate</td>
<td>• Set up websites</td>
</tr>
<tr>
<td>• Modify documents</td>
<td>• Create online training applications</td>
</tr>
<tr>
<td>• Manage versions</td>
<td>• Keep personal data</td>
</tr>
<tr>
<td>• Share knowledge</td>
<td>• Manage databases</td>
</tr>
<tr>
<td>• Retain/dispose</td>
<td>• Share media</td>
</tr>
<tr>
<td></td>
<td>• Store renditions</td>
</tr>
<tr>
<td></td>
<td>• Archive PSTs</td>
</tr>
</tbody>
</table>
Duplicate factoids

• Duplicates around 20-30% of storage
  • Much of that is also eTrash
  • Average of 3 documents per duplicate set
  • 15% of the duplicates cannot be deleted
  • 85% of the duplicates were created by humans, 15% by systems
• Storage level deduplication does not benefit productivity, backups, eDiscovery, retention management
Range of actions to take

<= Stuff I want to migrate and control in ECRM
<= Stuff I want to schedule the retention of
<= Stuff I want to keep
<= Stuff I need to produce
<= Stuff I can’t inventory, but can’t get rid of
<= Stuff I need to ask someone’s permission
<= Stuff somebody else owns that is abandoned
<= Stuff I need to create a policy before I delete
<= Stuff I can delete now
Act

- Report
- Export for review
- Export for migration to new email archive or ECM (SharePoint)
- Export for legal review
- Delete
- Copy
- Move (Hide)
- Encrypt
- Verify
- Redact and Produce
Recent Information Management News

- The risk is real
- The value of compliance is immense

Facebook could face up to $1.6 billion in fines over data breach as regulators eye formal probe

Facebook disclosed that it had discovered a security bug that allowed hackers to access information to around 50 million accounts.

The Irish Data Protection Commission said it was looking into whether to open a formal investigation into Facebook.

The social network could be fined a maximum of $1.63 billion if it is found to have breached the General Data Protection Regulation (GDPR) in the European Union.

MD Anderson slapped with $4.3M penalty for HIPAA violations

An HHS administrative law judge upheld an HHS Office for Civil Rights finding against the University of Texas MD Anderson Cancer Center in Houston to pay $4.3 million in civil penalties for HIPAA violations related to the organization’s encryption policies. HHS confirmed June 10.

Ticketmaster Data Breach Could Equal GDPR Fines in Millions

It was announced this past weekend that Ticketmaster suffered a security breach, which reportedly included personal and payment information for 40,000 users in the United Kingdom. The company can now face millions in fines under the GDPR laws.

New Jersey fines Virtua Medical $418,000 for HIPAA breach

Healthcare IT News

GLOBAL EDITION

TOPICS

Verizon might buy Charter, which just bought Time Warner Cable

Technology

By Chris Baraniuk
Technology reporter

A Canadian analytics firm that worked for Vote Leave has received the UK’s first formal notice under a key data law, the UK’s data protection watchdog has confirmed.

AggregateIQ (AIQ) was accused of processing people’s data “for purposes which they would not have expected”.

The firm has appealed against the notice, which was issued by the UK’s Information Commissioner’s Office.

The ICO said that although the data was gathered before 25 May, when the GDPR regulations came into effect, it was concerned about the “continued retention and processing” of data after that date. Thus, it said, meant GDPR applied to AIQ’s handling of that information.

Earlier this year it was linked to UK data firm Cambridge Analytica by whistleblower Chris Wylie, who alleged that Cambridge Analytica improperly acquired Facebook data belonging to 50 million people via a third party.

Cambridge Analytica has been credited with helping Donald Trump win the US presidential election in 2016.

It was paid nearly £2.7m ($3.6m) by Vote Leave to target ads at prospective voters during the Brexit referendum campaign.

It was also used by pro-Brexit youth group Leave.EU.
What is compliance?

• “Compliance” is conformity to some criteria

• Sources of compliance criteria
  - Laws & regulation (GDPR, CCPA, 10 other states, many other countries)
  - Industry standards (ANSI, ISO)
  - Company policy (RIM, E-mail, BYOD, IT Security)
  - Business operational requirements
  - Best practices

Regulation components often include:

• Practices
• Reporting
• Recordkeeping
• Access
• Security
• Breach response
• Data mapping
Privacy Regulations

https://www.dlapiperdataprotection.com
Transform Content

• Change the environment to make regulatory response, protection and general IG (RM, Litigation) better, faster, easier
  • Tag
  • Move PII and other sensitive content to specific locations or into security models
  • Audit trail
  • Organize by function

• Also consider encryption, changing security, writing metadata to source content for protection
Drain the privacy lake

• Data Minimization
  • Don’t collect data
  • Only collect what data you need
  • Get rid of what metadata you don’t need
  • Anonymize what you can
  • Eliminate privacy content when no longer needed
  • Eliminate all content according to RS
Audit Policy

• Validation
  • Have data sources changed?
  • Have regulations changed?
  • Have processes changed?

• Gather and communicate metrics
  • How frequently and current?
  • Overlap with other regulations?

• Continual improvement

• Leverage ability to update critical data sources
Plan Incident Response

- Identify **responsible** parties and subject matter experts – map discussions and topics to groups
  - Define what “jewels” you are protecting, RULES and where that data resides, DATA MAP to link back to investigation
  - Identify responsible parties, entities tracked

- Take advantage of data map details to identify who creates, receives, accesses critical info to assign responsibility
Protect Data

• Based on the data map – what types of data need to be protected in what ways.
  • Segmentation
  • Technology sets
  • Groups and security on folders
  • Data minimization
  • Data flow patterns – why and where important data is being left unprotected areas

• Leverage endpoint monitoring to track attempts to access data inappropriately

Appropriate Data $\leq$ Appropriate Access
# Protecting Data

 PHF=Personal, Health, Financial

<table>
<thead>
<tr>
<th>Data</th>
<th>is breached by</th>
<th>and protected by</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHF, Passwords</td>
<td>POS transaction malware, Web page malware redirects</td>
<td>Network Security, Activity Monitoring</td>
</tr>
<tr>
<td>PHF</td>
<td>Logging into databases and redirecting data</td>
<td>Network security, database security</td>
</tr>
<tr>
<td>PHF</td>
<td>Logging into servers and copying (or encrypting) data tables</td>
<td>Segmentation, data minimization, security models</td>
</tr>
<tr>
<td>Intellectual property, PHF, Reputational</td>
<td>Logging into file shares and copying (or encrypting) documents/files</td>
<td>Segmentation, data minimization, security models, auditing</td>
</tr>
<tr>
<td>PHF, IP, Passwords</td>
<td>Losing laptops</td>
<td>Audits, process mapping, policy, encryption</td>
</tr>
<tr>
<td>PHF, IP, Reputational</td>
<td>Emailing data</td>
<td>DLP</td>
</tr>
</tbody>
</table>
Breach Monitoring and Response

• Monitoring current activities to determine
  • Who or what is doing suspicious things

• Investigation past activities to determine
  • What was breached?
  • What personal data was in that breach?
  • Who had access?
  • Who did it?
  • How to keep it from happening again?
Access Request

• Validate request
  • Find the data
    • Review and validate
    • Redact and produce
Benefits

Strategic
Visibility into highest value corporate assets
Optimize litigation, investigation and mitigation efforts
Alignment with strategic goals, IG stakeholders and external regulators

Operational
Focused on real tasks rather than voluminous access requests
Ability to focus on real threats and issues from attack
Consistency and repeatability

Tactical
90% reduction in content needing review
10x speed increase in review process
Compressed response time
Where to Start

1. Content Assessment
   • Identify and validate compliance data types
   • Low impact investment and starter pack
   • Begin mapping
   • Prioritize strategy elements

2. Incident Plan

3. Pen-Test

4. Expand when ready