Data Mapping & Metadata – What Can They Do for My IG Program?

Mary Sherwin, FAI, CRM, IGP, CIPP/US
David Smythe, MBA, IGP
Agenda & Takeaways

**Agenda**

- Why is this important to us?
- Data Mapping Overview & Illustration
- Business Scenarios
- Value, Challenges, Risks
- Privacy Data Mapping
- Tool Examples

**Takeaways**

- How this pertains to you
- Using the terms correctly
- Understand relevant scenarios
- Identify the problems to solve
- Understand perspectives – IG, IT
- Gain exposure to tools
Why is this Important to Us?

- Applying retention
- Managing inventories, data sets
- Privacy
- Data protection
- Identifying systems of record
- IT Asset Inventory
- Records Management System
In computing and data management, data mapping is the process of creating data element mappings between two distinct data models. Data mapping is used as a first step for a wide variety of data integration tasks. For example, a company that would like to transmit and receive purchases and invoices with other companies might use data mapping to create data maps from a company’s data to standardized ANSI ASC X12 messages for items such as purchase orders and invoices.
IT Data Mapping
Data Flow Mapping
**Metadata**

**Data: Business Information**
- Customer
- Products
- Suppliers
- Employees
- According Data

**Metadata: Rules and Pointers**

<table>
<thead>
<tr>
<th>Metadata Type</th>
<th>Points to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record metadata</td>
<td>Physical Table</td>
</tr>
<tr>
<td>Tablemap</td>
<td>Table join</td>
</tr>
<tr>
<td>Datamap</td>
<td>Columns in Tablemap</td>
</tr>
<tr>
<td>Filter</td>
<td>Where clause on tablemap/datamap</td>
</tr>
<tr>
<td>Tree metadata</td>
<td>Physical Tree</td>
</tr>
</tbody>
</table>

**Simple Examples**
- Title
- Creator
- Subject
- Description
- Publisher
- Date
- Type
- Format
- Source
- Language
- Rights

**Enterprise Suite**
**Business Scenarios**

Electronic Records – Metadata Management

Data Management – Data Mapping
Business Scenarios

Privacy – Metadata and Data Mapping

Application Catalog – Metadata
Mapping and Metadata Value

- **Data Mapping**
  - Data transformation or data mediation between a data source and a destination
  - Consolidation of multiple databases into a single database
  - Identifying redundant columns of data for consolidation or elimination

- **Metadata Management – Key to Information Lifecycle Management**
  - What type of information a repository contains (e.g., documents, images, structured database elements)
  - Where your data is (repositories and jurisdiction)
  - Where records and “special information” reside


- Reduce data loss
- Support incident response
- Ease data Migration and automation
- Improve data quality
- Enhance traceability, lineage
Data Mapping Challenges and Risks

- Not all data sources are known or mapped
- Tools are available to scan for data, but they are not nirvana
- Understanding what your objectives are is critical
- Some data is not readily available
- Critical starting factors
Data Mapping Scenario – Mergers & Acquisitions

- **Current State**
  - Multiple Systems with Similar / Overlapping Functions
  - Multiple Record Sources
  - Privacy Data throughout Systems and Records

- **Desired Future State**
  - Streamlined / Integrated Systems
  - One Record Source
  - All Privacy Data Identified and Mapped
  - Duplicate Records and Information Eliminated
Create the Data Map (1)

- Interview Business Units
  - Identify Systems and Records Repositories
  - What is the Frequency of Use?
  - Who accesses the System?
  - Is Retention Applied to the Data?
  - Document the Work and Data Flows

- Does the Data:
  - Contain “Special” Information?
  - Need Privacy Classification?
  - Need Security Classification?
Create the Data Map (2) – Partnerships

- Interview IT
  - Identify Systems and Records Repositories
  - Where is the System and Data Physically Located? (CCPA and GDPR Requirements)
  - Who Accesses the System?
  - What is the Work and Data Flow?
  - Is Retention Applied to the Data?
  - System Back-Up and Restore Processes?

- Interview Information Privacy
- Interview Information Security
## Components of Data Map

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Volumes</td>
<td>- Shared drives; SharePoint; Content Management Systems</td>
</tr>
<tr>
<td></td>
<td>- Messaging or social apps; email; email archives</td>
</tr>
<tr>
<td></td>
<td>- Structured data</td>
</tr>
<tr>
<td>Application</td>
<td>- Application version</td>
</tr>
<tr>
<td></td>
<td>- Data sources; data feeds</td>
</tr>
<tr>
<td></td>
<td>- Location; jurisdiction and organizational unit use</td>
</tr>
<tr>
<td></td>
<td>- Regulatory requirements (e.g., PII, hold orders); records classes</td>
</tr>
<tr>
<td></td>
<td>- Client requirements (e.g., ROT remediation, deduplication, tagging)</td>
</tr>
<tr>
<td>Unstructured Data</td>
<td>- Word Processing, Spreadsheets, Presentations</td>
</tr>
<tr>
<td></td>
<td>- Images (OCR requirements)</td>
</tr>
<tr>
<td>Structured Data</td>
<td>- System of record?</td>
</tr>
<tr>
<td></td>
<td>- Data archiving</td>
</tr>
<tr>
<td></td>
<td>- COTS; Platform; on premises; cloud-based</td>
</tr>
<tr>
<td></td>
<td>- Accessibility to scan databases</td>
</tr>
<tr>
<td></td>
<td>- Existence or need for encryption, masking, tokenization</td>
</tr>
</tbody>
</table>
## IT Data Mapping Tools

<table>
<thead>
<tr>
<th>Data Mapping Tools</th>
<th>Best For</th>
<th>Capabilities or Connectors</th>
<th>Data Formats Supported</th>
<th>Use Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>OneTrust</td>
<td>Privacy, security and third-party risk management.</td>
<td>Data flow mapping, visualization, compliance, technology, privacy services.</td>
<td>Databases, file systems, Hadoop.</td>
<td>Privacy</td>
</tr>
<tr>
<td>TrustArc</td>
<td>Privacy compliance and risk management</td>
<td>Data inventory/mapping, integrations, privacy services.</td>
<td>Databases, file systems, Hadoop.</td>
<td>Privacy</td>
</tr>
<tr>
<td>Global IDs</td>
<td>Data profiling, integration and migration suite.</td>
<td>Data discovery, support for data lineage.</td>
<td>Databases, file systems, Hadoop.</td>
<td>Data governance, discovery, rationalization, lineage</td>
</tr>
<tr>
<td>Pentaho</td>
<td>Its drag and drop functionality.</td>
<td>It provides support to NoSQL, Hadoop, object store, &amp; analytics database distributions.</td>
<td>Data ingestion from Apache Kafka.</td>
<td>Real-time data ingestion, Analytics, Operationalize data science.</td>
</tr>
<tr>
<td>Talend</td>
<td>It is best as a data integration tool.</td>
<td>Databases, flat files, and cloud-based applications.</td>
<td>XML &amp; XHTML etc.</td>
<td>--</td>
</tr>
<tr>
<td>Informatica</td>
<td>Power Center tools.</td>
<td>AWS Redshift, Azure SQL Data Warehouse, and Snowflake</td>
<td>XML, JSON, AVRO, PDF FILES, Microsoft Word, Excel.</td>
<td>Data Transformation, B2B Data Exchange,</td>
</tr>
</tbody>
</table>
Summary

- Data mapping and metadata management are related, but are very different in the eyes of IG and IT.

- There are opportunities for IG/RIM professionals in a number of ways, including:
  - Apply Retention to Electronic Records
  - Manage Your Inventories
  - Working with IT, Privacy and Security
  - Manage the Information and Data Lifecycle
  - Engage in Digital Transformation efforts

- Become More Integral in Information Governance Initiatives
Questions / Discussion

Thank You

Mary Sherwin, FAI, IGP, CRM, CIPP/US

mwsinny0593@gmail.com
https://www.linkedin.com/in/marysherwin

David Smythe, MBA, IGP

david.smythe@ironmountain.com
https://www.linkedin.com/in/david-smythe