

MAPPING



YOUR WAY TO COMPLIANCE WITH A DATA ATLAS

Attention on data maps has been brought to the forefront of organizations' senior management, thanks to new mandates in litigation discovery, but with collaboration among key stakeholders, a data map can be developed into a data *atlas*, providing a "total information systems overview" that will be an even more valuable tool for effective information governance.

Wayne Wong

The unprecedented proliferation of digital data has increased the pressure for organizations to operate effectively in an increasingly complex information governance environment. The records and information management (RIM) profession is in an era of increased regulatory pressure for retention management, audits, security, and privacy compliance for personally identifiable and personal health information.

Adding to these pressures are the 2006 amendments to the *Federal Rules of Civil Procedure* (FRCP), which explicitly clarify that electronically stored information (ESI) is discoverable and change previous document discovery strategy by mandating early pre-trial conferences of opposing attorneys to identify, disclose, and agree upon data specific to relevant information.

These issues, as well as many corporate initiatives – to successfully implement an enterprise-wide archiving solution, for example – demand that organizations have detailed knowledge of where information lives, what life cycle should be enforced, and keywords and metadata to be indexed with the full text to facilitate searches and retrieval beyond content text searching. A good data map provides this.

The amended FRCP refers to a data map as a prudent practice to meet today's e-discovery requirements. Indeed, walking into an FRCP Rule 26(f) "meet and confer"

session with a good data map is a proactive way to limit the scope of the discovery. This is significant given that several studies confirm that discovery comprises 50% (on average) of the cost of litigation. Hence, it becomes critical for organizations to proactively understand and manage their data in order to be able to efficiently and confidently respond to an incident or litigation within the newly mandated timeframe.

Understanding Data Maps

A *data map* is a catalog of an organization's ESI by category, location, and custodian or steward, including how it is stored, its accessibility, and associated retention policies and procedures. *Unstructured data*, the user-created documents (e.g., word processing documents) with no predefined structure or central repository, is particularly troublesome to manage.

To be effective, a data map must function beyond an asset inventory and serve to answer the following questions:

- What specific information exists?
- What is the volume of data?
- What period of time does the data cover?
- Who and what does the data involve?
- Where is the data located?
- What form is the data in?

As it turns out, the concept is straightforward, but the execution can be complex, and, currently, there is no standard depiction of a data map. Furthermore, to remain useful, the data map must be kept “evergreen” and perpetual rather than treated as a one-time initiative.

There are shortcomings to a data map. The term has a very different meaning and context in the IT world, as identified by Wikipedia: “Data mapping is the process of creating data element mappings between two distinct data models [defined below]... used as a first step for a wide variety of data integration tasks ...” It is used to map elements in two fields, for example, between two databases for migration or integration purposes.

Data models, as noted by Wikipedia, are primarily used by IT professionals to “precisely explain a subset of real information to improve communication within the organization and thereby lead to a more flexible and stable application environment ... they typically do not describe unstructured data, such as word processing documents, email messages, pictures, digital audio, and video.”

The term “map” also generally connotes, and sets an expectation of, a graphical depiction. But as attractive as that notion may be, each graphical map on paper is limited to a single resolution and detail level making it difficult to portray every data detail (e.g., life cycle or metadata) as may be required by every particular need and concern (e.g. litigation, investigation, or retention management needs).

Graphical depictions are also very expensive to maintain and keep up-to-date. Perhaps someday, someone will create a visual application that is the equivalent of “Google Earth” where an individual can visually start at the global enterprise level and be able to dynamically and graphically “zoom” into particular data repositories or into sets of backup tapes to find the details about what data exists for the organization at any point in time. In the interim, though, organizations can begin to get this type of overview by going beyond the data map to create what would more accurately be referred to as a data atlas.

Creating a Data Atlas

A *data atlas* is a collection or compendium of maps, charts, lists, and tables with supplementary illustrations and analyses that describe the information and the infrastructure and systems used to host the information, resulting in a “total information systems overview.” Graphical depictions are excellent for certain aspects of the data story, and an atlas allows for the inclusion of maps and charts without excluding lists, spreadsheets, databases, and other analyses.

Creating and implementing a data atlas is a multi-disciplinary process, involving RIM, IT, legal, and subject matter experts (SMEs). Lines-of-business teams often include compliance, IT security, human resources, finance, research and development, and other departments that maintain records. The need for collaboration, a unified agenda, and a common understanding of key terms, such as “record” and “policy,” cannot be emphasized enough. It is helpful to establish and emphasize the key role RIM professionals play in this process, as they have the broadest view of the need for managing records and information.

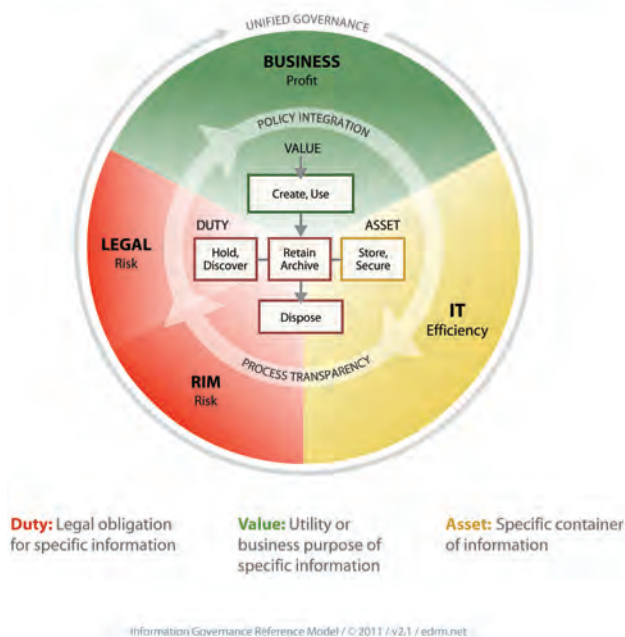
Utilizing the new information governance reference model (IGRM) by EDRM (*edrm.net*) can help facilitate communication and clarify respective roles among this collaborative group. The IGRM model (see image at left) is not prescriptive in nature; rather, it provides a reference that will promote cross-functional dialogue and collaboration. It provides common language and reference for discussion and decision making.

Central to the reference model are the respective roles, where:

- Legal and RIM are focused on risk and are responsible for the *duty* of information, or the legal and regulatory obligation of specific information.
- IT is focused on efficiency and is responsible for the *asset* (stewardship) of information, or the specific containers of information.
- SMEs are focused on profit and are responsible for the *value* of information, or the utility or business purpose of specific information.

Information Governance Reference Model (IGRM)

Linking duty + value to information asset = efficient, effective management



Step 1: Organize a Multi-Disciplinary Team

The first step in creating a data atlas is to assess and organize the multi-disciplinary stakeholders and conduct scoping and charter meetings with them. (As mentioned earlier, at a minimum, the participants should include RIM, IT, legal, and line-of-business leaders and SMEs.) To establish a data atlas is to establish a corporate culture change; hence, enlisting a senior management sponsor will increase the likelihood of enterprise-wide adoption and support.

The various disciplines should be tasked with surveying and gathering current regulatory requirements,

policies, procedures, inventories, lists, and diagrams. Ideally, a centralized repository and collaboration space should be created to house the following business units' information:

RIM:

- Latest, up-to-date data retention policies, retention schedules, and procedures
- ARMA International's Generally Accepted Record-keeping Principles® (GARP®) (More information can be found at www.arma.org/garp.)

IT:

- Published policies (e.g., technology use policies and procedures, existing asset and application inventories, and current IT network and server topology maps)
- Format, lifecycle, and locations of e-mails and other messaging systems
- Desktop and laptop computer data storage policies at the organizational and individual levels
- Policies and processes for data backups
- Policies for repurposing a computer after an employee has left the organization
- Disaster recovery plan and protocol
- Older tape sets, which may present additional obstacles as the original tape reading equipment and versions of the authoring software may no longer be available

Legal:

- Obligations under the FRCP
- Other jurisdictional obligations that may apply (e.g., state and municipal)
- Ongoing litigation matters
- Custodians
- Active litigation holds
- Other obligations based on how highly regulated the organization's industry is (e.g., it may be regulated by the Financial Industry Regulatory Authority, Securities and Exchange Commission Rule 17a, the Health Insurance Portability and Accountability Act, the Health Information Technology for Economic and Clinical Health Act, or the Payment Card Industry data security standards)

Line-of-Business Leaders and SMEs:

- Other obligations and special data handling that might be missed (e.g., applicable sections of the IRS code or Food and Drug Administration regulations)
- Specific procedures for data handling and filing when making the connection between the logical storage locations where they maintain data and the physical network locations familiar to the IT team

How to Represent the Data Atlas

Historically, even large organizations used spreadsheets to represent data maps. The spreadsheet paradigm of rows of data types and columns of attributes is intuitive and generally the most easily understood representation model. However, as litigation and organizations have grown more complex and the amount of electronic data has grown exponentially, the spreadsheet model becomes more and more cumbersome.

A good data atlas goes beyond the data map by describing not only the data content, but the system metadata (e.g., author, creation, modification dates, and infrastructure and systems) used to host the information resulting in a "total information systems overview." In practical experience, the most flexible and useful solutions are query-able relational databases.

Fortunately, the spreadsheet model of labeled rows and columns logically upgrade into database models of records and fields in a straightforward fashion. Database tables describing information can be related to other tables and applications that document regulations and policies that apply to it or that dictate special handling requirements. Databases also:

- Facilitate "automated" updating by integrating with other existing processes, such as purchasing or an incident/support ticketing system
- Generate result sets that can then be rendered or depicted in various forms, such as lists, tables, charts, or graphical maps – essentially a virtual data atlas.

Step 2: Identify Information in All Repositories

Identify and conduct interviews with information stewards and custodians to identify the information within various data repositories, data systems, and main applications. The interviews should:

- Shed light on what kind of information is stored within each repository, as well as how to access each repository
- Gather lifecycle information to determine how far back the data covers
- Identify any discrepancies between what is actually kept and what is mandated by policy; this should be recorded and flagged for possible remediation later.

Step 3: Use Findings to Create Data Atlas

The results of the interviews – along with other pre-existing information contained within other existing enterprise systems – will provide the information needed to create the various pieces that make up the data atlas, which may include such things as flow charts, a master spreadsheet, databases, or all of these. (See sidebar “How to Represent the Data Atlas” on page 24.)

Step 4: Test the Results

Test the data atlas on a real litigation matter or a hypothetical investigation. RIM should challenge the data atlas to see if it is able to provide all the information needed to automate a retention rule or policy for a hypothetical retention archive. RIM should know where the data is and in what form, and RIM should have access to the metadata to automate the identification and retention classification. RIM will then be able to judge whether the data map created is applicable to real-world data requests.

Implementing the Atlas

Consider a Consultant

Partner with an experienced consultant to help keep the implementation focused, to keep it on track, and to help determine if it makes sense to implement the data atlas in phases. Due to the magnitude of the undertaking, it is generally most cost-effective to do the majority of the work internally and utilize the consultant for guidance and high-level project management assistance, rather than pay him or her hourly to do the work.

Consider Vendor Solutions

There are proprietary, dedicated solutions offered by several vendors to implement data mapping. Currently, the leading packages provide more workflow assistance than provide a graphical map, but workflow management is very helpful to ensure the resulting data atlas remains evergreen.

Maintaining the Atlas

Whether you use a propriety product or not, plan for integrating the updating of the data atlas into existing processes, such as procurement, help desk support, IT change management or asset management programs, and other infrastructure support functions, to ensure the data atlas does not become obsolete and lose value.

If a new storage unit is added or new litigation arises, and a litigation hold must be invoked, the data atlas must have the infrastructure and custodian contact information needed, and there must be a consistent, repeatable, and documented process in place to keep the data atlas current. Best practice also mandates development of a governance process to manage monitoring, periodic reviews and updates, and remediation.

The business case for dedicated software platforms that allow multiple authorized parties to interact with and update the data map is clear. Today’s RIM protocols simply require richer, more dynamic, more detailed, and more accurate atlases of enterprise data. **END**

Wayne Wong can be contacted at lit.prep@hotmail.com. See his bio on page 47.