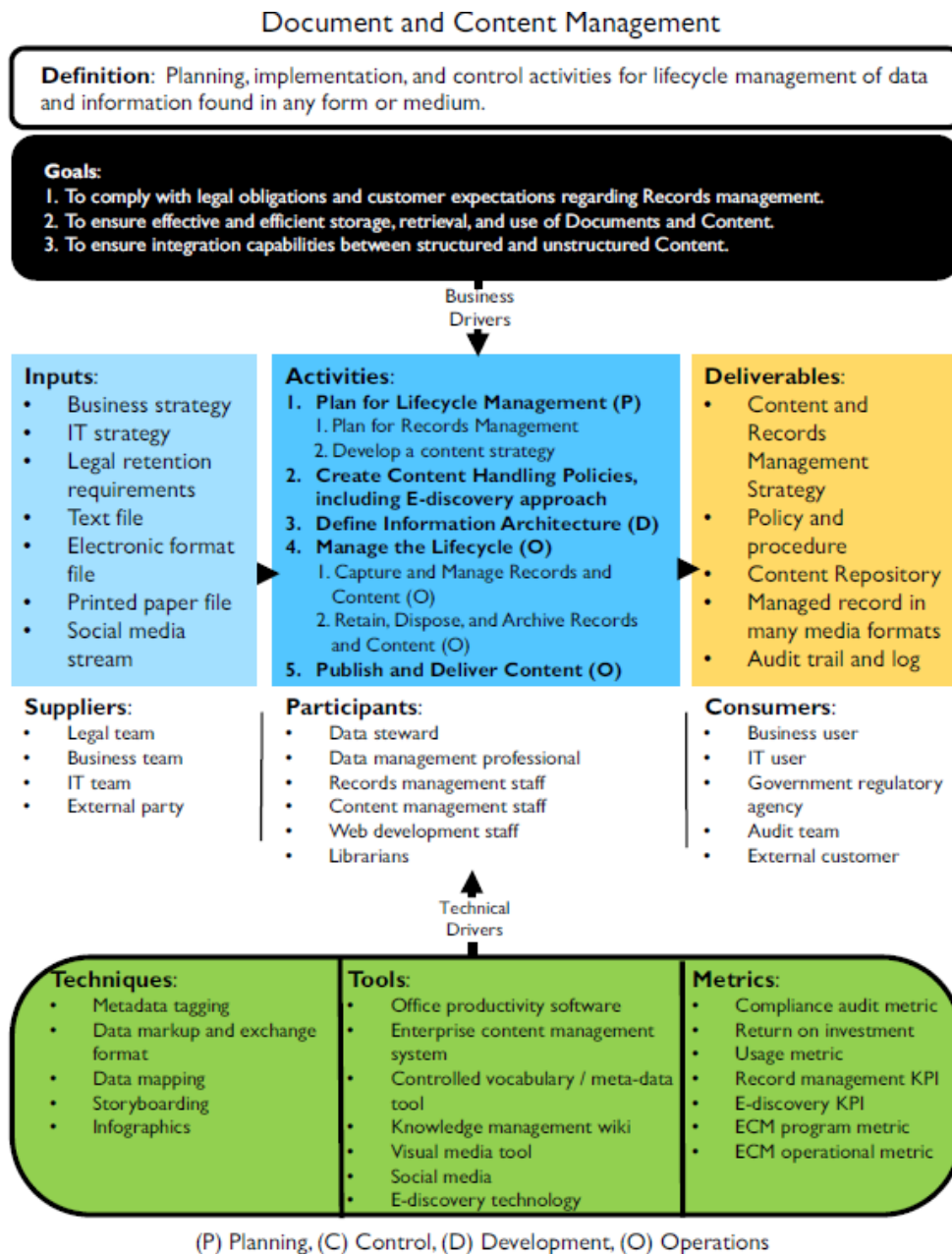# Document and Content Management

1 Introduction

Document and Content Management entails controlling the capture, storage, access and use of data and information stored outside relational databases. Records include both paper documents and ESI (Electronically Stored Information).

Document and Content Management

Definition: Planning, implementation, and control activities for lifecycle management of data and information found in any form or medium.

Goals:
1. To comply with legal obligations and customer expectations regarding Records management.
2. To ensure effective and efficient storage, retrieval, and use of Documents and Content.
3. To ensure integration capabilities between structured and unstructured Content.

Business Drivers

Inputs:
- Business strategy
- IT strategy
- Legal retention requirements
- Text file
- Electronic format file
- Printed paper file
- Social media stream

Activities:
1. **Plan for Lifecycle Management (P)**
   1. Plan for Records Management
   2. Develop a content strategy
2. **Create Content Handling Policies, including E-discovery approach**
3. **Define Information Architecture (D)**
4. **Manage the Lifecycle (O)**
   1. Capture and Manage Records and Content (O)
   2. Retain, Dispose, and Archive Records and Content (O)
5. **Publish and Deliver Content (O)**

Deliverables:
- Content and Records Management Strategy
- Policy and procedure
- Content Repository
- Managed record in many media formats
- Audit trail and log

Suppliers:
- Legal team
- Business team
- IT team
- External party

Participants:
- Data steward
- Data management professional
- Records management staff
- Content management staff
- Web development staff
- Librarians

Consumers:
- Business user
- IT user
- Government regulatory agency
- Audit team
- External customer

Technical Drivers

Techniques:
- Metadata tagging
- Data markup and exchange format
- Data mapping
- Storyboarding
- Infographics

Tools:
- Office productivity software
- Enterprise content management system
- Controlled vocabulary / meta-data tool
- Knowledge management wiki
- Visual media tool
- Social media
- E-discovery technology

Metrics:
- Compliance audit metric
- Return on investment
- Usage metric
- Record management KPI
- E-discovery KPI
- ECM program metric
- ECM operational metric

(P) Planning, (C) Control, (D) Development, (O) Operations

1.1 Business Drivers

- **Regulatory compliance:** Respond efficiently and consistently
- **The ability to respond to litigation**
- **e-discovery:** The process of finding electronic records that might serve as evidence in litigation. Includes email, chats, websites, raw application data and Metadata. Huge volume of ESI.
- **Business continuity requirements**

- **Efficiency**: Use technology to streamline processes, manage workflow, eliminate repetitive manual tasks and enable collaboration
- **Customer satisfaction**: Well organised and searchable websites

1.2 Goals and Principles

Goals:
1. To comply with legal obligations and customer expectations regarding Records management.
2. To ensure effective and efficient storage, retrieval, and use of Documents and Content.
3. To ensure integration capabilities between structured and unstructured Content.

Goals of implementing best practices around Document and Content Management:

- Ensuring effective and efficient retrieval and use of data in unstructured formats
- Ensuring integration capability between structured and unstructured data
- Complying with legal obligations and customer expectations

Management of Document and Content follow these guiding principles:

- Everyone in the organisation must create, use, retrieve and dispose of records in accordance with the established policies and procedures. Train everyone.
- Experts in the handling of records and content should be fully engaged in policy and planning

In 2009 ARMA International published GARP (Generally Acceptable Recordkeeping Principles® how business records must be maintained: Details on ARMA website

- **Principle of Accountability:**
  - Senior executive appropriates individuals, adopts policies and processes and ensures auditability
- **Principle of Integrity:**
  - Guarantee of authenticity and reliability
- **Principle of Protection:**
  - Personal information is protected
- **Principle of Compliance:**
  - Program should comply with laws, binding authorities and the policies of the organisation
- **Principle of Availability:**
  - Information is managed in a manner that ensures timely, efficient and accurate retrieval of information
- **Principle of Retention:**
  - Retain information for an appropriate time taking legal and regulatory requirements into consideration
- **Principle of Disposition:**
  - Provide secure and appropriate disposition of information in accordance with operational, legal, regulatory and fiscal requirements
- **Principle of Transparency:**
  - Policies, processes, activities, governance program documented and understood by all staff and appropriate interested parties

1.3 Essential Concepts

1.3.1 Content

Document is to content what a bucket is to water: a container. **Content** refers to the data and information inside the file, document or website. Content has a lifecycle and can become an official **record**. Official records are treated differently.

- **Content Management:** the processes, techniques and technologies for organising, categorising and structuring information resources so that they can be stored, published and reused.
- **Content Metadata:** Essential for managing unstructured data and "Big Data":
  - **Format:** often dictates the method to access
  - **Search-ability:** search tools already exist
  - **Self-documentation:** Metadata self-documenting as in file systems
  - **Existing patterns:** can they be adopted
  - **Content subjects:** The things people are likely to be looking for
  - **Requirements:** Is a content tagging tool necessary?
- **Content Modelling:** The process of converting logical content concepts into content types, attributes and data types with relationships using Metadata Management and Data Modelling techniques.
  - Information product level e.g. website
  - Component level
- **Content Delivery Methods:** Convert to XML to enable reuse over many channels. Types of delivery systems:
  - **Push:** Users choose type of content to be delivered on a schedule.
  - **Pull:** Users pull content through the Internet
  - **Interactive:** e.g. electronic point of sale. High volumes of real-time data

1.3.2 Controlled Vocabularies

A defined list of explicitly allowed terms to index, categorise, tag, sort and retrieve content through browsing and sorting. Controlled vocabularies should be aligned with the entity names and definitions in the enterprise conceptual model. May be considered Reference Data and Metadata.

- **Vocabulary Management:** The function of defining, sourcing, importing and maintaining any given vocabulary.
  - ANSI/NISO Z39.19-2005 definition: Way to improve the effectiveness of information storage and retrieval systems, web navigation systems and other environments that seek to both identify and locate desired content via a description using language.
  - Purpose of vocabulary control is consistency in the description of content objects to facilitate retrieval
- **Vocabulary Views and Micro-Controlled Vocabulary:**
  - a vocabulary view is a subset of a vocabulary relevant to a group of users
  - A micro-controlled vocabulary is a vocabulary view containing highly specialised terms
- **Term and Pick Lists:** lists in applications that do not contain relationships
- **Term Management:** One or more words designating a concept (ANSI/NISO Z39.19-2005). Terms should be specified and managed through governance processes. Three types of term relationships:
  - **Equivalent term relationship:** Leads to one or more terms to use

- o **Hierarchical relationship:** depicts broader (general) to narrower (specific) or whole-part relationships.
  - o **Related term relationship:** a term that is associatively but not hierarchically linked to another term in a controlled vocabulary.
- **Synonym rings and Authority lists:**
  - o **Synonym ring:** a set of terms with roughly equivalent meaning. A search on one reveals content on the others
  - o **Authority list:** a controlled vocabulary of descriptive terms designed to facilitate retrieval of information within a specific domain
- **Taxonomies:** a naming structure containing a controlled vocabulary used for outlining topics and enabling navigation and search systems. Can be a data model.
  - o **Flat taxonomy:** no relationships among the set of controlled categories
  - o **Hierarchical taxonomy:** Tree structure where nodes are related by a rule. Going up expands the category and going down refines.
  - o **Polyhierarchy:** Tree structure with more than one rule. Multiple parents. Paths are complex
  - o **Facet taxonomy:** Star structure. each node is associated to a central node e.g. Metadata where each attribute is a facet of a content object
  - o **Network taxonomy:** Both hierarchical and facet structures. e.g. a thesaurus
- **Classification Schemes and Tagging:** Codes that represent controlled vocabulary. Folksonomy: names are obtained though social tagging
- **Thesauri:** A Thesaurus is a controlled vocabulary used for content retrieval. Provides information about each term and its relationship to other terms.
- **Ontology:**
  - o a type of taxonomy that represents a set of concepts and their relationships within a domain.
  - o Semantic Web.
  - o Developed using ontology languages such as RDFS (Resource Description Framework Schema)
  - o Describe classes (concepts), Individuals (instances) attributes, relations and events.
  - o two **differences** between taxonomy and ontology:
    - **Taxonomy:** data content classifications for a given concept area.
    - **Ontology:** Content concepts are mixed and identified through Metadata.
    - **Taxonomy:** closed-world assumption, only what is known is defined.
    - **Ontology:** open-world assumption, possible relationships are inferred

1.3.3 Documents and Records

Documents are electronic or paper objects which communicate information and knowledge.

Records are a subset of documents which provide evidence of actions taken and decisions made in accordance with procedures. Can be created manually or automatically by equipment.

- **Document Management:** Processes, techniques and procedures for controlling and organising documents and records throughout their lifecycle.
  - o **Inventory:** Identification of existing and newly created documents and records
  - o **Policy:** Creation, approval and enforcement of documents and records policies
  - o **Classification**
  - o **Storage:** Short and long term physical storage

- o **Retrieval and circulation:** allow access on accordance with policies, security and legal standards
- o **Preservation and disposal:** Archiving and destroying documents and records according to organisational needs, statutes and regulations.
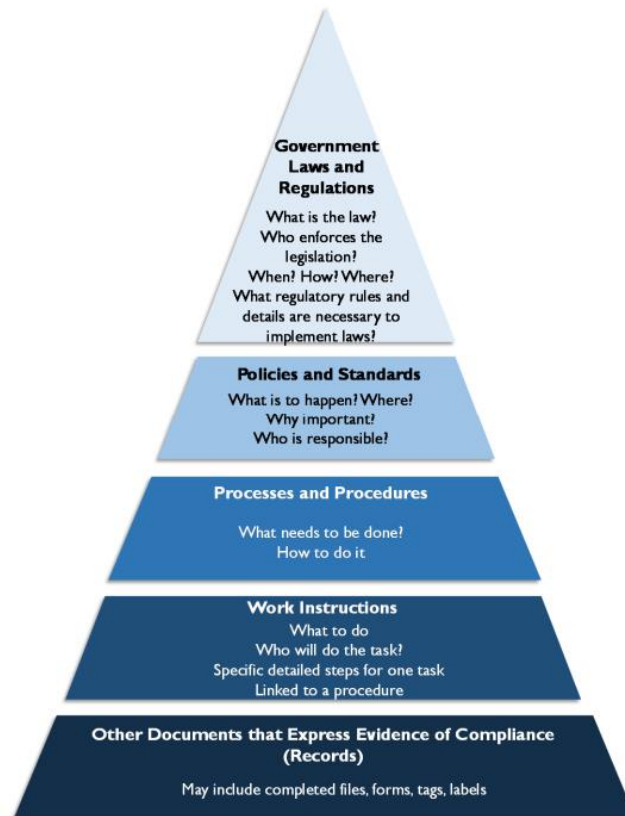

Figure 72 Document Hierarchy based on ISO 9001-4.2

- **Records Management:**
  - o Managing records through the full lifecycle:
    - Record creation or receipt
    - processing
    - distribution
    - organisation
    - retrieval
    - disposition
  - o Records can be physical, electronic, web content, documents on all kinds of media, data in any kind of database.
  - o A **Vital Record**: required by an organisation to resume after a disaster
  - o Trustworthy records have signatures for regulatory compliance and integrity.
  - o Characteristics of well-prepared records:
    - **Content**: Accurate, complete, truthful
    - **Context:** Collect Metadata
    - **Timeliness:** created promptly after the event
    - **Permanency:** Records may never be changed
    - **Structure:** Content should be clear and legible

- **Digital Asset Management:** Document management for rich media documents (videos, logos, photographs etc.)

1.3.4   Data Map

A data map is an inventory of all ESI (Electronically Stored Information) data sources, applications and IT environments.

1.3.5   E-*D*iscovery

Discovery is the pre-trial phase of a lawsuit where both parties request information from each other to find the facts of the case. Rules were amended in 2006 to include ESI (Electronically Stored Information)

- Electronic documents have Metadata which plays a part in evidence.
- Legal requirements:
  - e-discovery
  - data and records retention practices
  - Legal hold notification (LHN) process
  - Legally defensible disposition practices
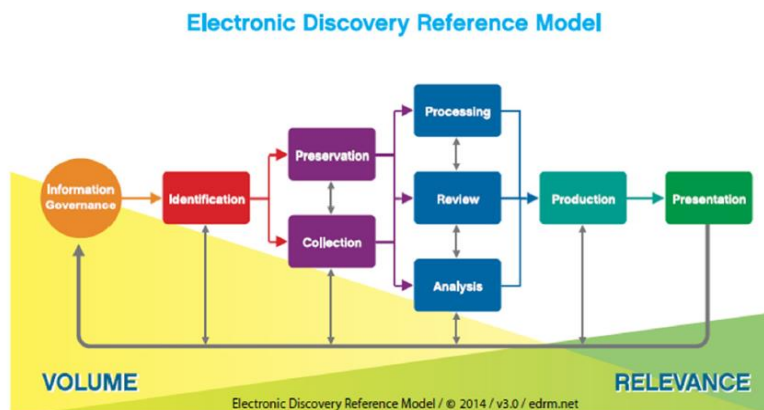
EDRM is a standards and guidelines organisation for e-discovery. The EDRM model assumes information governance is in place. Eight e-discovery phases:

- **Identification:**
  - **Early Case Assessment:** Legal case assessed for descriptive information or Metadata
  - **Early Data Assessment:** Type and location of data relevant to the case is assessed.
    - Identify policies for retention or destruction of relevant data so ESI can be preserved
    - Interviews with involved personnel to get pertinent information
    - Involved personnel need to understand the case background, legal hold and their roles
- **Preservation:** Identified potentially relevant data is put on hold to avoid destruction
- **Collection:** Acquisition and transfer information to legal counsel in a legally defensible manner
- **Processing:** Data de-duplicated, searched and analysed to determine which moves to review
- **Review:** Documents identified to be presented on request
- **Analysis:** Understand the content to formulate a legal strategy
- **Production:** Information turned over to opposing counsel based on agreed to specifications
- **Presentation:** Displaying the ESI at depositions, hearings and trials

1.3.6   Information Architecture

Information architecture is the process of creating structure for a body of information or content. Identifies links and relationships between documents and content. Components:

- Controlled vocabularies
- Taxonomies and ontologies
- Navigation maps
- Metadata maps
- Search functionality specifications
- Use cases
- User flows

The information architecture and the content strategy together describe WHAT content will be managed in the system. Design phases describe HOW the strategy will be implemented.

1.3.7   Search Engine

A search engine is software that searches for information based on terms and retrieves websites that have those terms within their content. e.g. Google.

1.3.8   Semantic Model

Semantic modelling is a type of modelling that describes a network of concepts and their relationships, and allows users to ask questions of the system in a non-technical way.

1.3.9   Semantic Search

A semantic search engine can use artificial intelligence to identify query matches based on keywords and their context. BI and analytics tool users often have semantics search requirements.

1.3.10   Unstructured Data (rather call it non-tabular or semi-structured data)

About 80% of all stored data is outside relational databases. Include word processing documents, emails, social media, chats, flat files, spreadsheets, XML files, transactional messages, reports, graphics, digital images, microfiche and video and audio recordings.

There is no data model to aid understanding. No tags. Not organised in rows and columns.

The fundamental principles of Data Management apply. Unstructured data requires data governance, architecture, security, Metadata and data quality.

1.3.11   Workflow

Manage content management through a workflow that ensures content is created on schedule and receives proper approvals. Should be automated through use of a content management system (CMS) to provide version control.

2   Activities

2.1   Plan for Lifecycle Management

Planning for document management:

- For the document's lifecycle
- Developing classification/indexing systems and taxonomies
- Creating policy specifically for records

Identify an organisational unit responsible for managing documents and records.

Chapter 9

2.1.1 Plan for records Management

Define what constitutes a record. The defining team for a functional area should include SMEs and records system people. Decisions:

- where to store and archive
- account for paper records as well as electronic unstructured and structured data

2.1.2 Develop a content Strategy

Defines how content will be prioritised, organised and accessed. Identify Content drivers (the reasons content is needed).

- inventory of current state
- gap assessment

2.1.3 Create Content Handling Policies

Policies help employees understand and comply with the requirements for document and records management.

Most document management programs have policies relating to:

- Scope and compliance with audits
- Identification and protection of vital records
- Purpose and schedule for retaining records
- How to respond to information hold orders (in the event of a lawsuit holding records beyond expiry of retention)
- Requirements for onsite and offsite storage of records
- Use and maintenance of hard drive and shared network drives
- email management, addressed from content management perspective
- Proper destruction methods
- **Social Media policies:** Does a social media post from an employee conducting business become a record?
- **Device Access Policies:** Content and records management functions need to work with user driven IT (BYOD bring-your-own-devices)
- **Handling Sensitive data:** Data Security and Data Governance establish confidentiality schemes and identify confidential and restricted assets. People who assemble content must apply these classifications.
- **Responding to litigation:** Prepare for the possibility of litigation through proactive e-discovery. Create an inventory of data sources and their associated risks.

2.1.4 Define Content Information Architecture

Users need to submit search to a system which contains unstructured data in a form understandable to that system. The inventory of structured and unstructured documents needs to be described or indexed for quick retrieval.

2.2 Manage the lifecycle

2.2.1 Capture records and content

Capturing data:

- Electronic content is in the form to be stored in electronic repositories
- Paper: scan and upload using an electronic signature

- Tag/index with the appropriate Metadata necessary for retrieval:
  - Document identifier
  - Date and time of capture
  - Title and author(s)

2.2.2 Manage versioning and control

ANSI Standard 859 has three levels based on criticality of the data:

- **Formal control:** Formal change control processes through change control authority
- **Revision control:** Notify stakeholders and increment versions when a change is required.
- **Custody control:** Least formal, requires safe storage and means of retrieval.

Table 15 Levels of Control for Documents per ANSI-859

Data Asset	Formal	Revision	Custody
Action item lists		x	
Agendas			x
Audit findings		x	x
Budgets	x		
DD 250s			x
Final Proposal			x
Financial data and reports	x	x	x
Human Resources data		x	
Meeting minutes			x
Meeting notices and attendance lists		x	x
Project plans (including data management and configuration management plans)	x		
Proposal (in process)		x	
Schedules	x		
Statements of Work	x		
Trade studies		x	
Training material	x	x	
Working papers			x

ANSI 859 recommends taking the following into account when determining control level for an asset:

- Cost of providing and updating the asset
- Project impact of changes in terms of cost
- Other consequences of change to the enterprise or project
- Need to reuse or earlier versions of the asset
- Maintenance of a history of change

2.2.3 Backup and recovery

Include the documents/record management system in the organisation's **corporate backup and recovery plan** and **Business Continuity Plan**. A **vital records program** allows the business to conduct business during a disaster and resume normal business afterwards. Vital records must be identified and plans for their protection and recovery must be developed and maintained.

2.2.4 Manage retention and disposal

A retentions and disposition policy defines (in accordance with legal and regulatory requirements):

- The timeframes during which documents for operational, legal, fiscal or historical value must be maintained

- When inactive documents can be transferred to a secondary storage facility
- Processes for compliance
- Methods and schedules for disposition of documents.

Records managers and information asset owners ensure privacy and data protections requirements and prevent identity theft.

Software considerations: may require a certain version to read. An upgrade can make documents unreadable or inaccessible.

Non-value-added information must be identified and removed as it wastes resources and can be discoverable in the event of litigation.

2.2.5 Audit Documents/Records

Table 16 Sample Audit Measures

Document / Records Management Component	Sample Audit Measure
Inventory	Each location in the inventory is uniquely identified.
Storage	Storage areas for physical documents / records have adequate space to accommodate growth.
Reliability and Accuracy	Spot checks are executed to confirm that the documents / records are an adequate reflection of what has been created or received.
Classification and Indexing Schemes	Metadata and document file plans are well described.
Access and Retrieval	End users find and retrieve critical information easily.
Retention Processes	The retention schedule is structured in a logical way either by department, functional or major organizational functions.
Disposition Methods	Documents / records are disposed of as recommended.
Security and Confidentiality	Breaches of document / record confidentiality and loss of documents / records are recorded as security incidents and managed appropriately.
Organizational understanding of documents / records management	Appropriate training is provided to stakeholders and staff as to the roles and responsibilities related to document / records management.

Steps of an audit:

- Define organisational drivers and stakeholders – the why
- Gather data on the process (the how), determine what to measure and what tools to use.
- Report the outcomes
- Develop an action plan of next steps and timeframes

2.3 Publish and deliver Content

- **Provide access, search and retrieval** once it is described by Metadata/key word tagging and classified within the appropriate information content architecture
- **Deliver through acceptable channels**: Users want to consume data on the device of their choice meaning the format may change. It may be difficult to bring it back to the original format e.g. HTML to originally structured.

3 Tools

3.1 Enterprise Content Management Systems

An ECM consists of a platform of core components or applications which can be in-house or in the cloud.

Reports can be delivered through common tools or a document management system interface allowing search, view, download and print. Report   management is

the ability to add, change or delete reports organised in folders. Reports retention can be set for automatic purge or archive.

3.1.1   Document Management

A **document management system** is an application used to track and store electronic documents and electronic images of paper documents. Provides storage, versioning, security, Metadata management, content indexing and retrieval.

Once a document is created within the document management system it is indexed with keywords so that it can be found. Metadata (Extracted automatically or added manually) is stored for each document. Bibliographic records of documents are descriptive structured data in MARC (Machine-Readable Cataloging) standard format.

A document repository enables check-in and check-out features, versioning, collaboration, comparison, archiving, status state, migration from one storage media to another and disposition.

May have a module that supports different types of workflows: Manual, rules-based or dynamic rules based on content.

Document management systems have a rights module to allocate permissions.

Specialised document management:

- **Digital asset management:** Document management systems may include management of audio, video and digital photographs
- **Image processing:** Captures, transforms and manages images of paper and electronic documents. Uses:
    - **Scanning:** digitising
    - **Optical Character Recognition (OCR):** Mechanical or electronic conversion of digitised print or handwriting to form recognised by software
    - **Intelligence character recognition (ICR):** more intelligent, convert cursive handwriting
    - **Form processing:** capture of printed forms. System recognises the format
    - **Other digitised images:** binary image file formats:
        - **vector**: mathematical formulae to create graphics. Can be resized without compromising resolution. File formats: .eps, .ai and .pdf
        - **raster**: bitmap of coloured pixels. .jpeg, .gif, .png, .tiff
- **Records management system:**
    - Automation of retention and disposition
    - e-discovery support
    - long term archiving to comply with legal and regulatory requirements
    - Vital records program for critical records

3.1.2   Content Management System

A **content management system** is used to collect, index and retrieve content, storing it either as components or whole documents while maintaining links between the components. It maintains content throughout the lifecycle

3.1.3   Content and Document workflow

Workflow tools support business processes, route content and documents, assign work tasks, track status and create audit trails.

3.2   Collaboration tools

For teams

3.3   Controlled vocabulary and Metadata tools

Office productivity software, Metadata repositories, BI tools and document and content management systems

3.4   Standard Markup and exchange formats

To allow computer systems to process unstructured data it must be reformatted.

- **XML:** Extensible Markup Language
    - Uses Metadata to describe content, structure and business rules of a document.
    - Translates structure of the document for data exchange
    - XML tags data elements to identify the meaning of data
    - Older markup methods are HTML and SGML
    - Need for XML-capable content management:
        - Integrating structured and unstructured data in a relational database
        - XML can build enterprise or corporate portals
        - XMP provides labelling and identification of unstructured data
- **JSON:** JavaScript Object Notation
    - Text format is language independent and easy to parse
    - two structures:
        - Objects: a collection of unordered name/value pairs
        - An array: an ordered list of values
    - Preferred format for web and NoSQL databases
- **RDF and related W3C specifications:**
    - **Resource Description Framework** is the standard for web data exchange
    - Stored in a **triplestore:** Subject (Resource) – predicate (property name) – Object (property value).
    - URI (Uniform Resource Identifier) describes subject-predicate-object.
    - URL (Uniform Resource Locator) is a form of URI
    - Semantic Web needs access to both data and relationships between data sets (Linked Data). This is provided by URIs
    - RDF uses XML as its encoding syntax
    - SKOS (Simple Knowledge Organisation System) is an RDF vocabulary
    - OWL (W3C Web Ontology Language) is a vocabulary extension of RDF used for publishing and sharing OWL documents on the web.
    - Big Data: If data is accessible using the RDF Triples model, SPARQL query language can be used to find patterns without predefining a schema.
- **Schema.org:** Labelling content with semantic markup

3.5   E-Discovery technology

Large volumes of data. Technology Assisted Review (TAR) is a process where a team can review selected documents and mark them relevant or not.   This is input to a predictive coding engine that reviews and sorts the remaining documents according to relevancy.

4   Techniques

- Litigation response playbook:

- Litigation Response data map

5 Implementation Guidelines

5.1 Readiness Assessment / Risk Assessment

Identify areas where content management improvement is needed. A Data Management Maturity Assessment Model can help.

Specific ECM critical success factors:

- Content audit and classification for existing content
- Appropriate information architecture
- Support of the content lifecycle
- Definitions of appropriate Metadata tags
- The ability to customise functions of the ECM solution

5.1.1 Records Management Maturity

ARMA Generally Accepted Recordkeeping Principles guides policy and practice for Records Management. ARMA Information Governance Maturity Model assesses the organisation's recordkeeping program:

- **Level 1 Sub-Standard:** Information governance and recordkeeping concerns not addressed or just minimally
- **Level 2 In Development:** Developing recognition information governance and recordkeeping can have an impact.
- **Level 3 Essential:** Minimum requirements to meet legal and regulatory requirements
- **Level 4 Proactive:** A proactive information governance
- **Level 5 Transformational**: Information governance is integrated into the corporate infrastructure and business processes

Analyse gaps and risks and their impact on the organisation. If an organisation does not inventory records it is already at risk.

5.1.2 E-Discovery Assessment

A readiness assessment examines and identifies improvements to the litigation process. A mature process will be documented, defensible and auditable and will specify:

- Clear roles and responsibilities
- preservation protocols
- data collection methodologies
- disclosure processes

5.2 Organisation and Cultural Change

ECM can lead to more tasks, such as scanning and defining Metadata.

Management of content and records may be siloed. Training may be necessary to enforce policies across the enterprise.

Content and records management should be elevated to a high priority activity organisationally, especially in highly regulated industries where the RIM function needs to be closely aligned with Corporate legal and e-discovery functions.

6 Document and Content Governance

6.1 Information Governance Frameworks

Documents, records and other unstructured content poses risk to an organisation. Drivers to managing the risk include:

- Legal and regulatory compliance
- Defensible disposition on records
- pro-active preparation for e-discovery
- Security of sensitive information
- Management of risks associated with email and Big Data

Principles of successful Information Governance programs:

- ARMA GARP® principles
- Assign executive responsibility for accountability
- Educate employees on information governance responsibilities
- Classify information correctly
- Ensure authenticity and integrity of information
- Determine that the official record is electronic unless specified directly
- Develop policies for alignment of business systems and third-parties to information governance standards
- Store, manage and make accessible, monitor and audit approve enterprise repositories and systems for records and content
- Secure confidential or personally identifiable information
- Control unnecessary growth of information
- Dispose information at the end of its lifecycle
- Comply with requests for information.
- Improve continuously

The Information Governance Reference model (IGRM) shows the relationship of Information Governance with other business functions. The outer ring shows the stakeholders who put policies, standards, processes, tools and infrastructure in place. The centre is a lifecycle diagram.

Chapter 9

Information Governance Reference Model (IGRM)
Linking duty + value to information asset = efficient, effective management
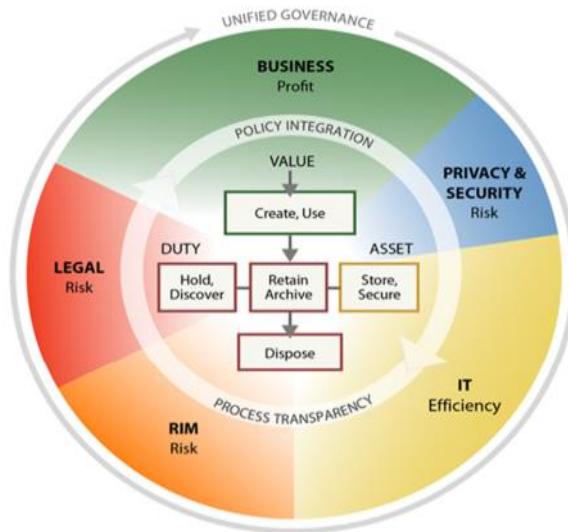
Duty: Legal obligation for specific information

Value: Utility or business purpose of specific information

Asset: Specific container of information

6.2 Proliferation of Information

The challenge of governance of unstructured data:

- Unstructured data grows much faster than structured data.
- Ownership can be difficult to ascertain
- Difficult to classify as business purpose cannot always be inferred
- Unstructured data without Metadata is a risk as it can be misinterpreted and may be mishandled

6.3 Govern for Quality Content

Requires effective partnership between data stewards and other data management professionals and records managers. Business data stewards can help define web portals, enterprise taxonomies, search engine indexes and content management issues.

Defining quality content requires understanding the context of its production and use:

- **Producers:** Who creates the content and why
- **Consumers:** Who uses the information and for what purpose
- **Timing:** When and how frequently the information is needed
- **Format:** Is a consistent format required? Are there unacceptable formats?
- **Delivery:** How the information is delivered, how it will be accessed and security to prevent inappropriate access

6.4 Metrics

KPIs at strategic and operational levels, especially if they measure lifecycle functions or risks.

6.4.1 Records Management

Strategic level:

Chapter 9

- Within regulatory requirements e.g. time taken to meet a requirement
- Governance e.g. compliance with policies

Operational level:

- Within areas of records management resources e.g. operational and capital costs
- Training
- Delivery of daily records management services and operations
- Integration of records management functions with other business functions

Criteria to measure the success of a records management system:

- Percentage of total documents and email per user identified as corporate records
- Percentage of these records put under records control
- Percentage of the total stored records that have proper retention rules applied.

Use ARMA's GARP principles and maturity model to guide definition of KPIs.

6.4.2   E-Discovery

KPIs of e-discovery:

- cost reduction
- efficiency of collecting documents ahead of time rather than reactively
- How quickly an organisation can implement a legal hold notification process (LHN)

6.4.3   ECM (Electronic Content Management)

KPIs measure tangible and intangible benefits to the organisation:

- Tangible:
    - Increased productivity
    - Cost reduction
    - Improved information quality
    - Improved compliance
- Intangible:
    - Improved collaboration
    - simplification of job routines and workflow