Applied A.I. Solutions

Foundations of Data Management

Professor
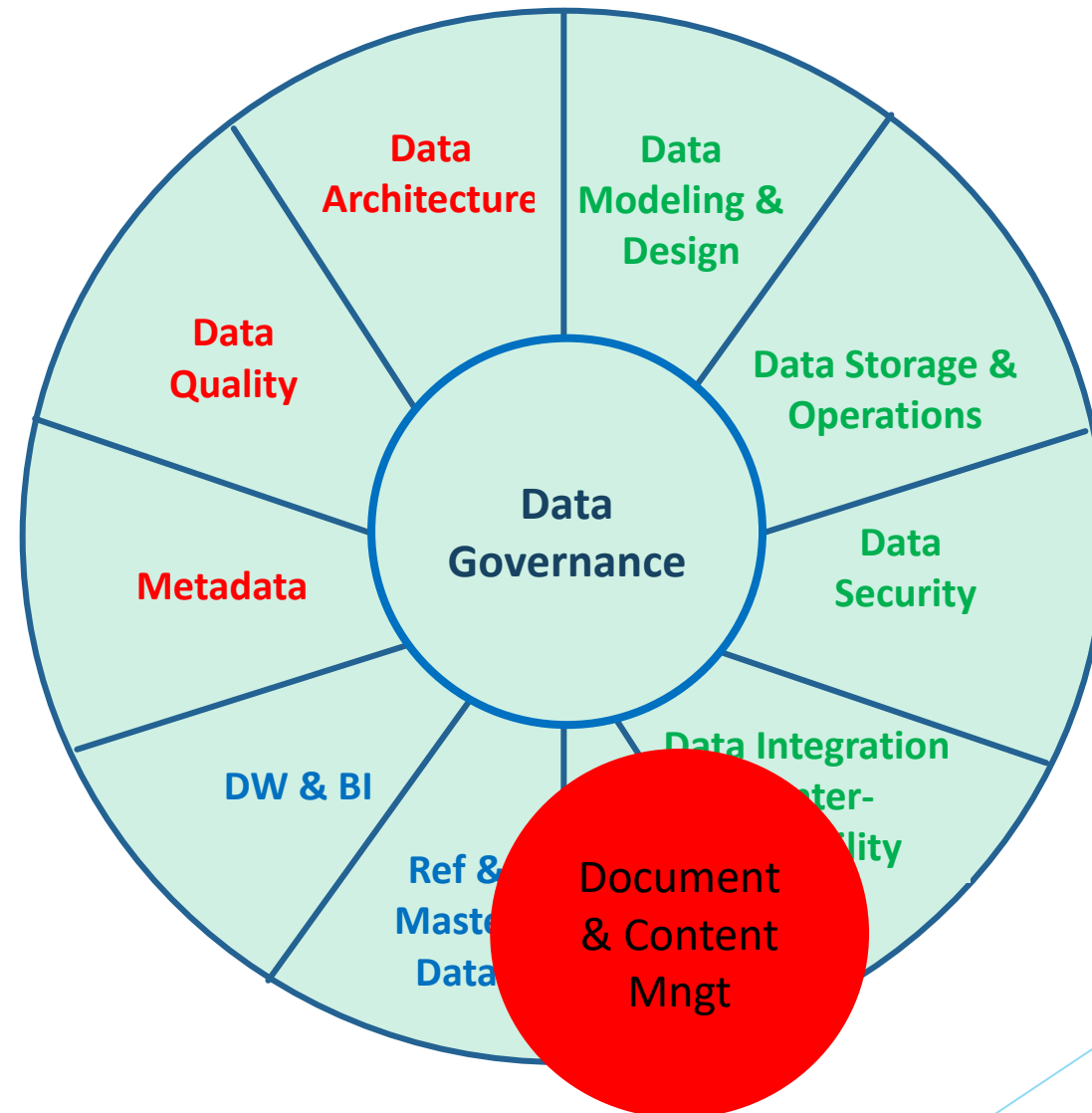# Daniel Vitaver-Bronstein, B.Sc., EMBA

daniel.vitaver-bronstein@georgebrown.ca

DOCUMENT AND CONTENT MANAGEMENT

The DAMA Wheel

1. INTRODUCTION

- ==Document and Content Management entails controlling the capture, storage, access, and use of data and information stored outside relational databases==

- Its focus is on maintaining the quality, security, integrity of and enabling ==access to documents and other unstructured information==

DCM Framework

Definition

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

DCM Framework
Goals

1. Comply with ==legal obligations, and customer expectations== regarding Records management

2. Ensure effective, efficient storage, retrieval, use of documents and content

3. Ensure ==integration capabilities between structured, unstructured content==

Business   Drivers

Principles:

1. <mark>Everyone must create, use, retrieve, and dispose of records in accordance with the established policies and procedures</mark>

2. Experts in handling of Records and Content should be fully engaged in <mark>policy</mark> and planning, and should comply with regulations and best practices

ARMA International   https://www.arma.org/

- A set of General Acceptable Recordkeeping Principles (ARMA International / GARP, 2009) describes how business Records should be maintained, and provides a Governance Framework with associate metrics. ARMA International   https://www.arma.org/

ARMA Principles of
- Accountability
- Integrity
- Protection
- Compliance
- Availability
- Retention
- Disposition
- Transparency

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

2. Create Content Handling Policies, including e-discovery

3. Define Information Architecture

4. Manage the Lifecycle

5. Publish and Deliver Content

Deliverables

1. Content and Records Management Strategy
2. Policy and Procedures
3. Content Repository
4. Managed Records in many media formats
5. Audit Trails and Logs

Suppliers

- Legal Team
- Business Team
- IT Team
- External Party

Participants

- Data Stewards
- Data management Professionals
- Records Management Staff
- Content Management Staff
- Web Development Staff
- Librarians

Consumers

- Business User
- IT User
- Government Regulatory Agencies
- Audit Team
- External Customer

Technical ⬆ **Drivers**

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

Tools
- Office productivity software
- ECMS
- Controlled vocabulary / Metadata tool
- KMS (wiki)
- Visual, Social media tools
- E-discovery technology

Metrics
- Compliance audit metric
- ROI
- Usage metric
- Record management KPI
- E-discovery KPI
- ECM program metric
- ECM operational metric

Drivers

- Regulatory compliance,

- Ability to respond to litigation and

- Ability to quickly respond to e-discovery requests (Big Data)

- Business continuity requirements

- Improve document management efficiencies

Essential Concepts

1. Content
2. Controlled Vocabularies
3. Documents and Records
4. Data Map
5. E-discovery
6. Information Architecture

7. Search Engine
8. Semantic Model
9. Semantic Search
10. Unstructured Data
11. Workflow

1. Content

- Content Management

- Content Metadata
  - Format
  - Search-ability
  - Self-documentation
  - Existing Patterns
  - Content Subjects
  - Requirements

- Content Modeling
  - Information level
  - Component level

- Content Delivery Methods
  - Push
  - Pull
  - Interactive

2. Controlled Vocabularies

- Vocabulary Management

- Vocabulary Views and Micro-controlled Vocabulary

- Term and Pick Lists

- Term Management

ANSI/NISO Z39.19-2005, American standard, guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies

2. Controlled Vocabularies – cont'd

- Synonym Rings and Authority Lists

- Taxonomies

- Classification Schemes and Tagging

- Thesauri

- Ontology

3. **Documents and Records**

- **Document Management** and Digital Asset Management
  - Inventory
  - Policy
  - Classification
  - Storage
  - Retrieval and Circulation
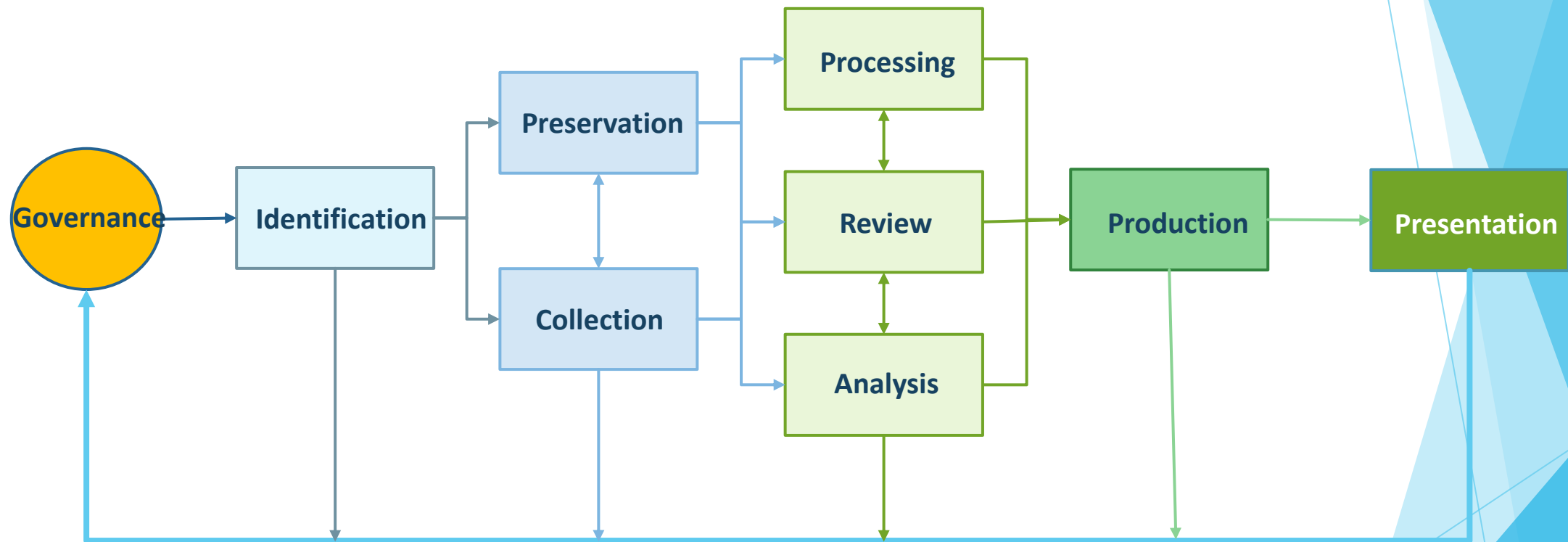 - Preservation and Disposal

- **Record Management**
  - Content
  - Context (Metadata)
  - Timeliness
  - Permanency
  - Structure

4. **Data Map**

References: ISO 9000/9001, ISO 15489, Canada's Bill 198

5. E-discovery

Reference Model (iterative)

6. **Information Architecture**

- Controlled Vocabularies, Taxonomies, ontologies
- Navigation, Metadata Maps
- Search Functionality Specifications
- Use Cases, User Flows

7. **Search Engine**

8. **Semantic Model**

9. **Semantic Search**

10. **Unstructured Data**

11. **Workflow**

2. ACTIVITIES

a) Plan for Lifecycle Management

- o Plan for Record Management

- o Develop a Content Strategy

- o Create Content Handling Policies

- o Define Content Information Architecture

b) Manage the Lifecycle

- o Capture Records and Content

- o Manage versioning and Control

- o Backup and Recovery

- o Manage Retention and Disposal

- o Audit Documents / Records

c) **Publish and Deliver Content**

- o Provide Access, Search, and Retrieval

- o Deliver Through Acceptable Channels

3. TOOLS

- Enterprise Content Management Systems

- Collaboration Tools

- Controlled Vocabulary and Metadata Tools

- Standard Markup and Exchange Formats (XML, JSON, RDF)

- E-discovery Technology

4. TECHNIQUES

- Litigation Response Playbook

  o defines the target environment for e-discovery

  o assess if gaps exist between current and target environments

- Litigation Response Data Map

  o defines a comprehensive catalog of information systems

5. IMPLEMENTATION GUIDELINES

- Readiness Assessment / Risk Assessment

  - Records Management Maturity
    (ARMA, GARP section 1.2)

  - E-discovery Assessment

- Organizational and Cultural Change

Standards
- DoD 5015.2 Electronic Records management Software Applications Design Criteria Standard
- ISO 16175, Principles and Functional Requirements for Records in Electronic Office Environments
- The Model Requirements for the Management of Electronic Records (MoReq2)
- The Records Management Services (RMS) specifications from the Object Management Group (OMG)

6. DCC GOVERNANCE

Drivers include:

- Legal and regulatory compliance

- Defensible disposition of records

- Proactive preparation for e-discovery

- Security of sensitive information

- Management of risk areas such as email and Big Data

- **Metrics**

  a) <mark>Records Management</mark>

    - % of total documents per user identified as corporate records

    - % of total stored records with proper retention rules applied

  b) <mark>E-discovery</mark>

    - Cost reduction

    - Efficiency gained in collecting information ahead of time