Applied A.I. Solutions

Foundations of Data Management

Professor

Daniel Vitaver-Bronstein, B.Sc., EMBA

daniel.vitaver-bronstein@georgebrown.ca

TOPICAL OUTLINE

Week 1	1.	Introduction to Data Management
	2.	Data Governance
	3.	Reference and Master Data Management,
	4.	Metadata Management
Week 2	1.	Data Architecture
	2.	Data Modeling and Design
	3.	Data Integration and Interoperability
	4.	Data Storage and Operations
Week 3	1.	Data Quality
	2.	Data Security
	3.	Document and Content Management
	4.	Data Privacy
	5.	Project Due, Live Presentation + Questionnaire

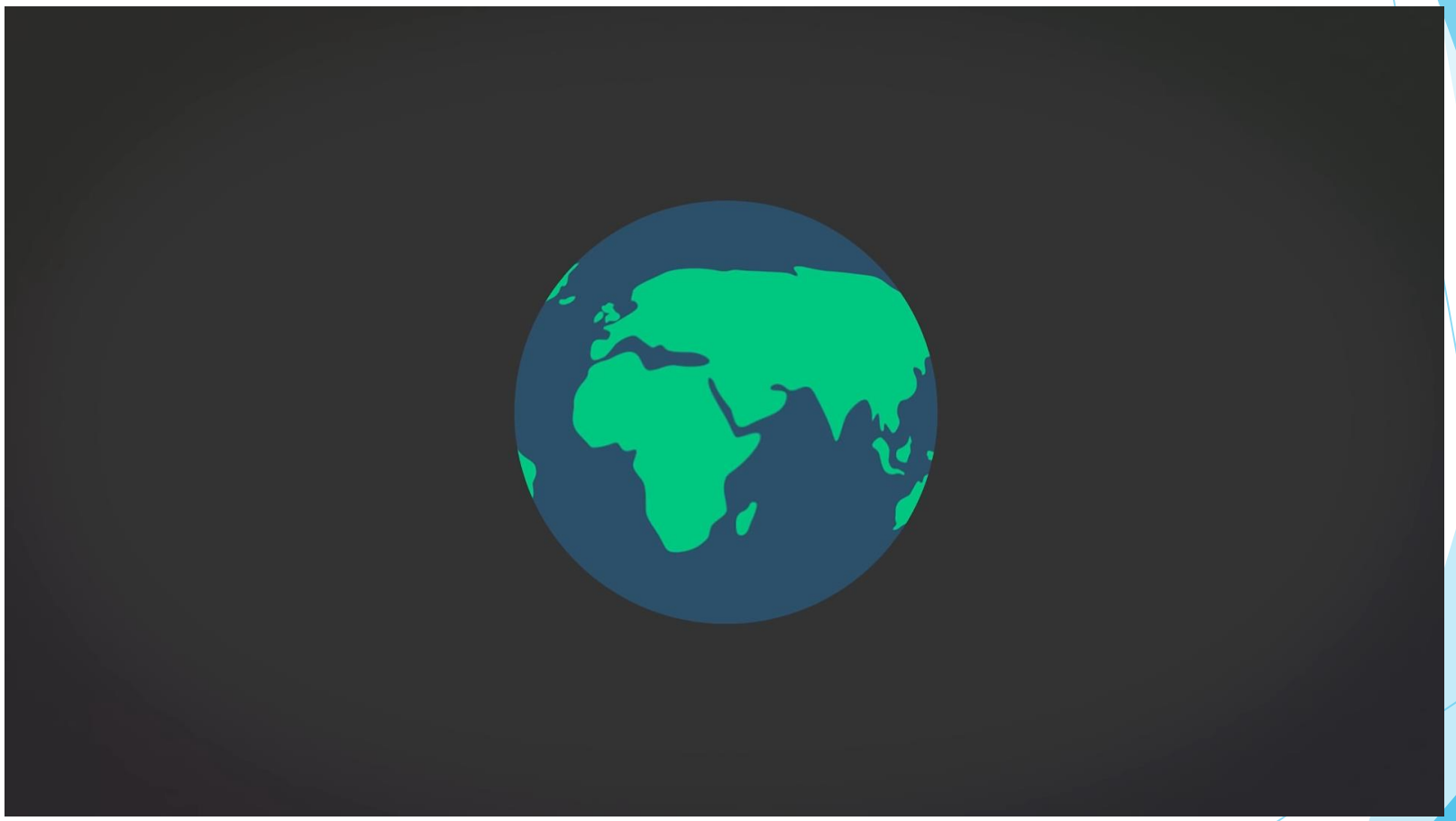
*This course can lead to a **CDMP** (Certified Data Management Professional) certification by DAMA*

Most Recommended Reading

DAMA International – DMBOK2 - Technics Publications,

Basking Ridge, New Jersey, USA

INTRODUCTION TO DATA MANAGEMENT



Source: Copyright © YouTube Video

Introduction

Data Management is the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles.

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data and Information - cont'd

- Understanding and supporting the information needs of the business and its stakeholders
- Capturing, storing, protecting, and ensuring the integrity of data assets
- Ensuring the quality of data and information

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data and Information - cont'd

- Ensuring the privacy and confidentiality of data
- Preventing unauthorized access, manipulation or use of data and information
- Ensuring data can be used effectively to add value to the business

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Essential Concepts - Data

- Data represents facts about the world (broad def.)
- Data are facts and statistics collected together for analysis (The New Oxford American Dictionary)
- Data is a mean of representation (Chisholm, 2010)
- Data is both an interpretation of the objects it represents and an object that must be interpreted (Sebastian-Coleman, 2013)

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

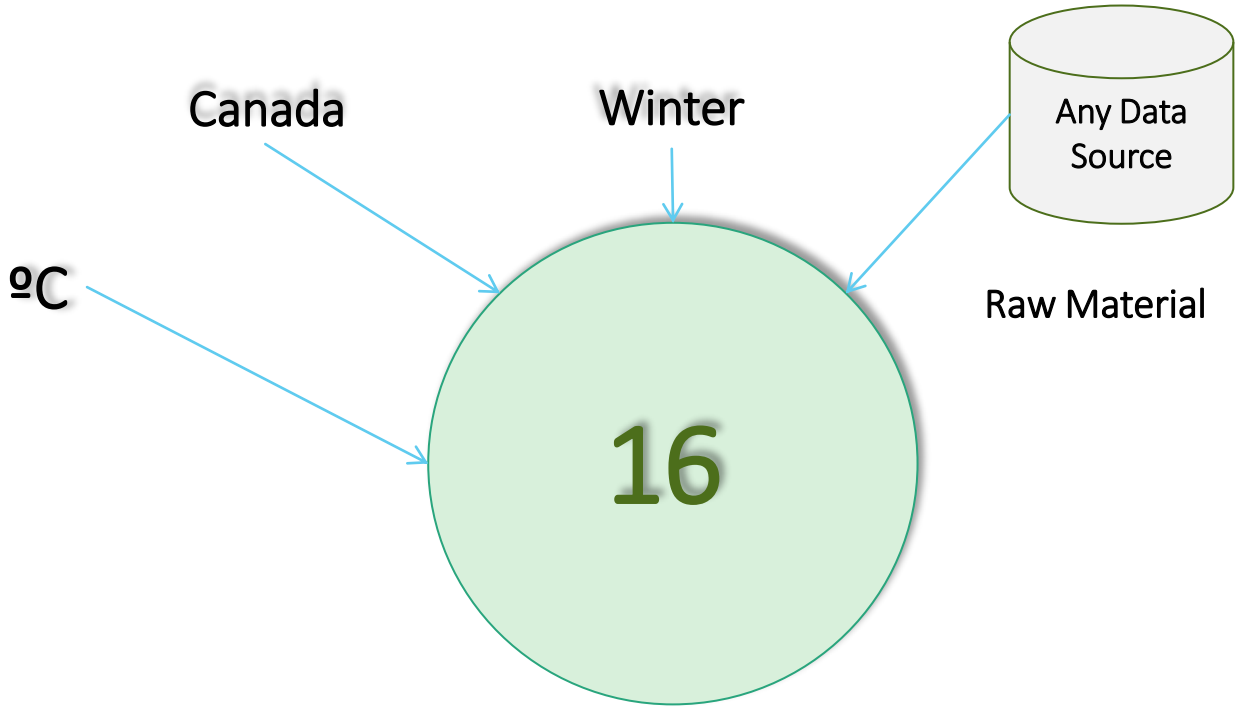
- **Data** are discrete, objective facts or observations, which are unorganised and unprocessed, and do not convey any specific meaning or value because it is without context and interpretation.[2]

- **Information** is data that have been contextualized and shaped into a form that is meaningful and adds value to the understanding of a subject and useful to human beings.[2]

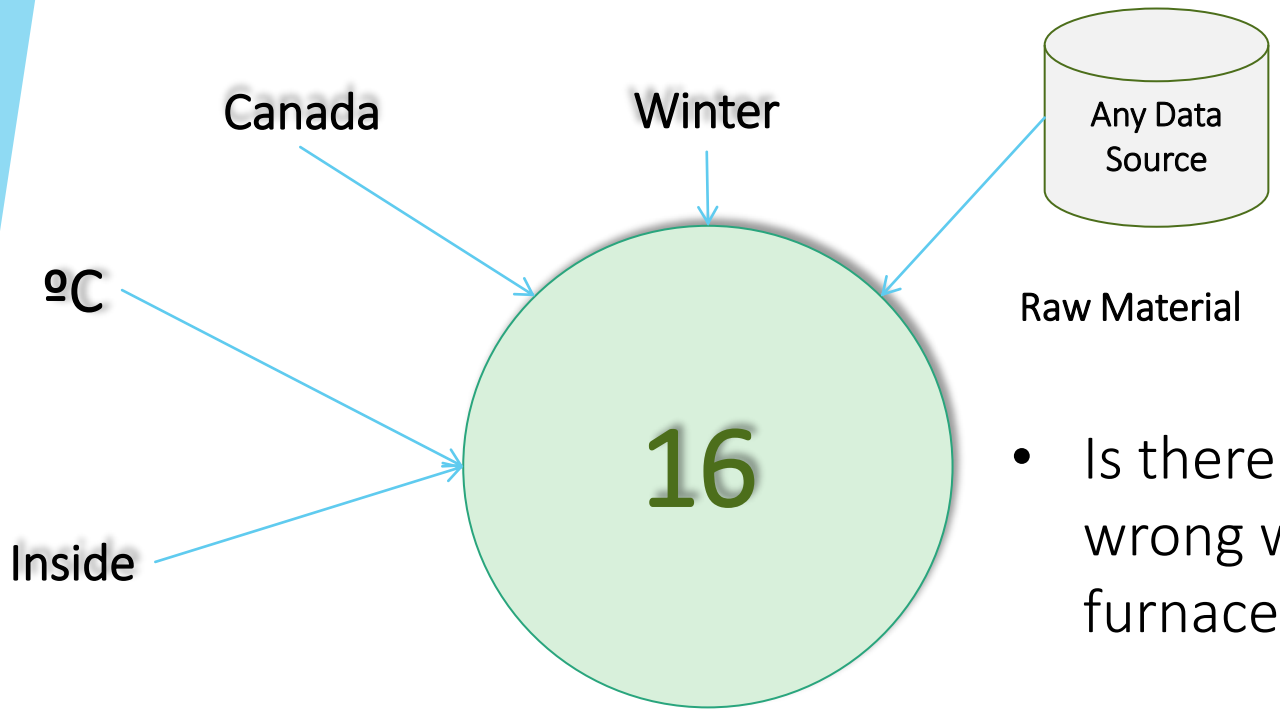
[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

[2] Laudon and Laudon, 2006, Turban et al., 2005, Boddy et al., 2005. (Laudon and Laudon, 2006, p. 13). (Awad and Ghaziri, 2004, Chaffey and Wood, 2005, Pearlson and Saunders, 2004, Bocij et al., 2003). (Awad and Ghaziri, 2004, p. 36). (Jessup and Valacich, 2003, Bocij et al., 2003, Groff and Jones, 2003). (Chaffey and Wood, 2005, p. 233).

What assumptions can you make …

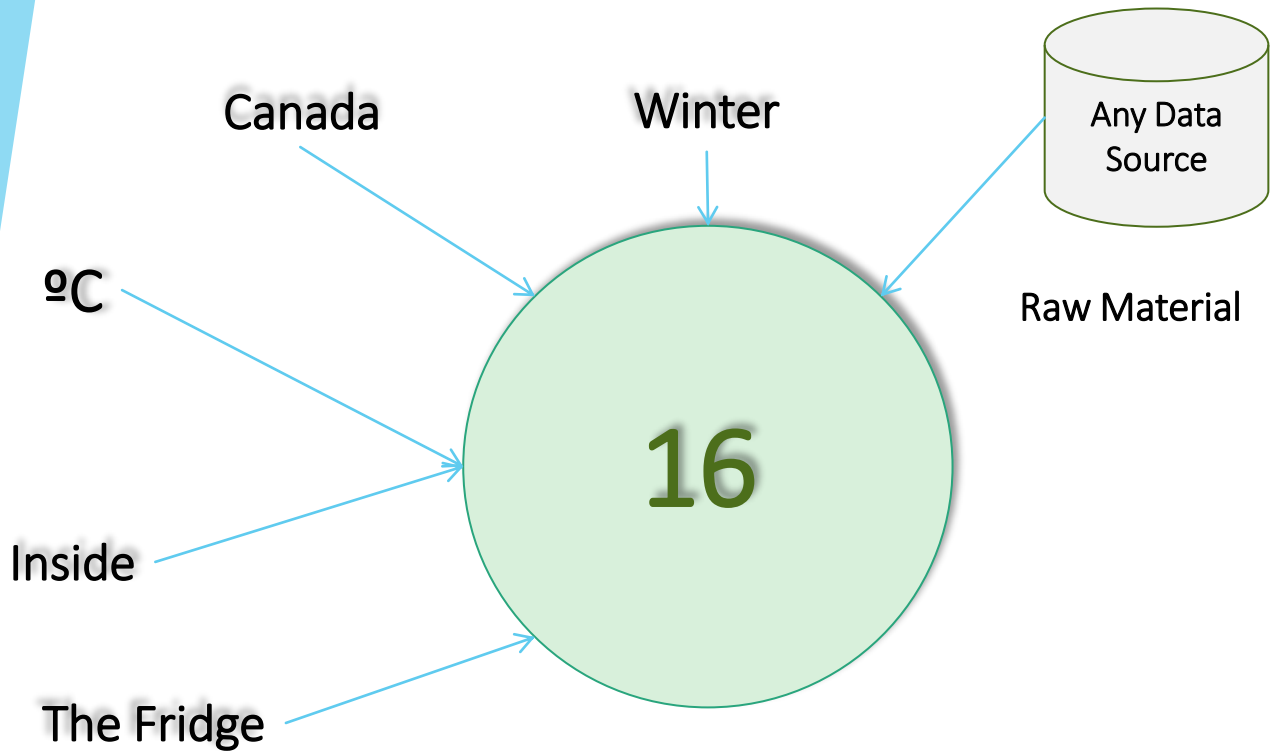


What assumptions can you make …

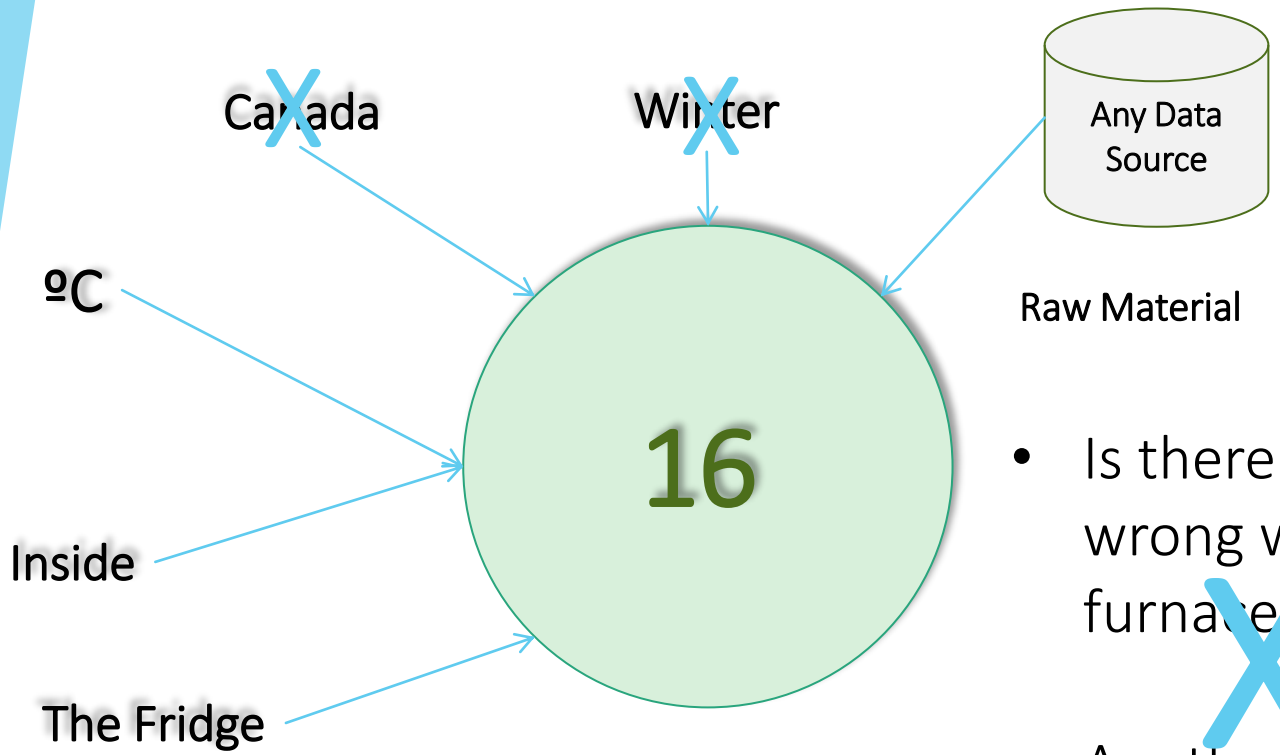


- Is there anything wrong with the furnace?
- Are there any windows open?

What assumptions can you make …



What assumptions can you make …



Context

- Is there anything wrong with the furnace?
- Are there any windows open?

Is there anything wrong with the fridge?

- Data does not simple exists. Data must be created
- It takes knowledge to create data in first place
- Data and Information are intertwined with and dependent of each other
- Data is a form of information and information is a form of data
- In DMBOK, the terms are used interchangeably

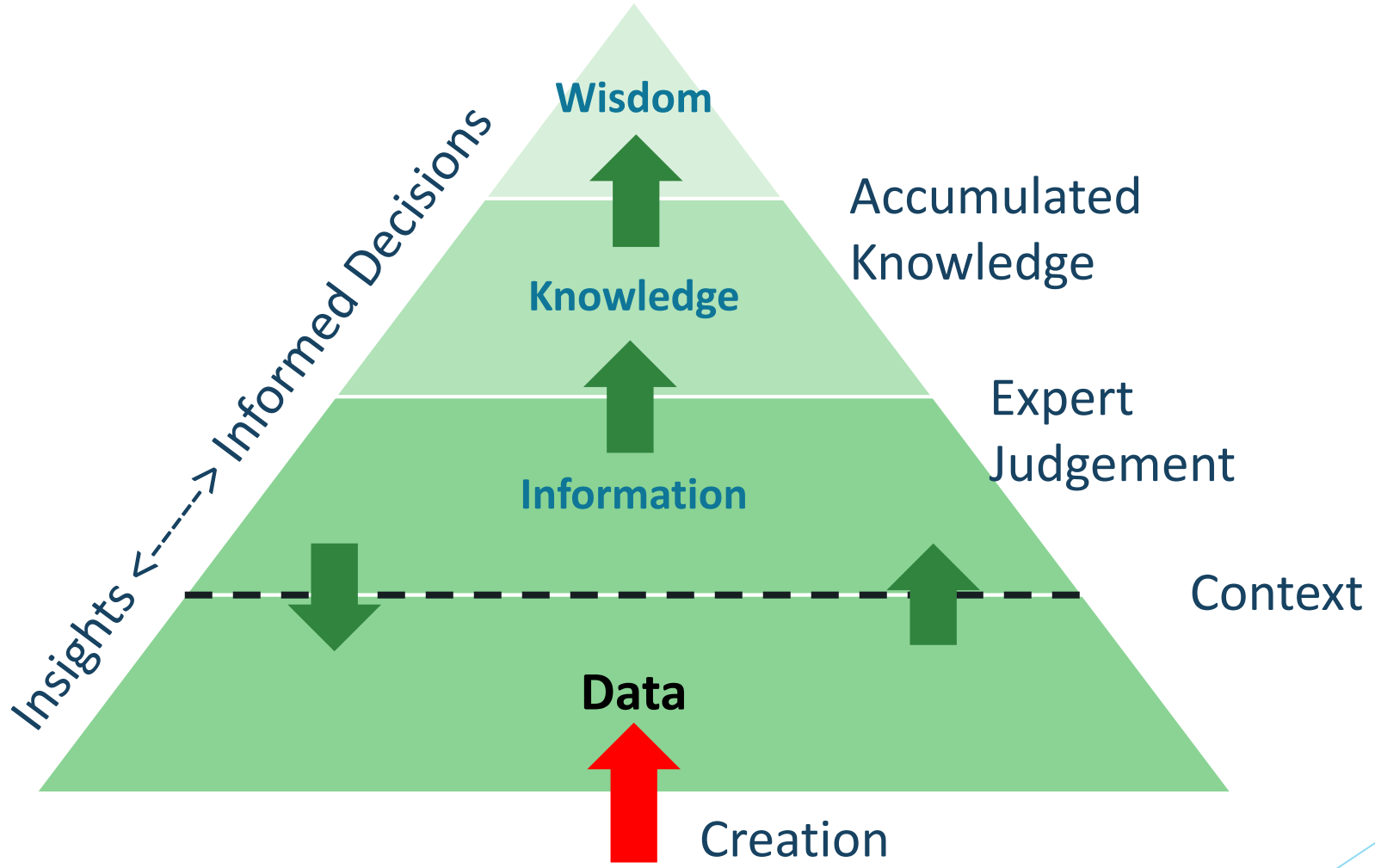
[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

- **Knowledge** is information to which is added expert opinion, skills, and experience, to result in a valuable asset which can be used for informed decision-making and performance improvement.[2]

- **Wisdom** is the accumulated knowledge, the highest level of abstraction and the ability to act critically or practically in any given situation.[2]

[2] Laudon and Laudon, 2006, Turban et al., 2005, Boddy et al., 2005. (Laudon and Laudon, 2006, p. 13). (Awad and Ghaziri, 2004, Chaffey and Wood, 2005, Pearlson and Saunders, 2004, Bocij et al., 2003). (Awad and Ghaziri, 2004, p. 36). (Jessup and Valacich, 2003, Bocij et al., 2003, Groff and Jones, 2003). (Chaffey and Wood, 2005, p. 233).

The DIKW Hierarchy



The DIKW Hierarchy - Rowley, 2007, p. 163

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Management Principles

1. Data is an asset with unique properties

2. Data value should be expressed in economic terms

3. Managing data means managing the quality of data

4. It takes metadata to manage data

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Management Principles – cont'd

5. It takes planning to manage data (architecture, processes, etc.)

6. Data management is cross-functional, requires skills, expertise and leadership vision, commitment, purpose

7. Data management requires an enterprise perspective and requires data governance program to be effective

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Management Principles – cont'd

8. Data management is lifecycle management

9. Different types of data have different lifecycle characteristics

10. Data management includes risk associated with data

11. Data management requirements must drive IT decisions – Data is dynamic, it constantly evolves

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Metadata

- Metadata is data about the asset (data)

- Metadata represents the context, information and knowledge about the data that we need to manage

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Metadata

- Metadata helps to interpret data in a meaningful way

- Metadata originates from a range of processes related to data creation, processing, and use, including architecture and governance

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Valuation

- **Value** is the difference between the cost of a product or service and the tangible and/or perceived benefit(s) derived form that product or service.

- Data is **unique** to each organization and thus data valuation is not standardized.

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Costs of

- Collecting, storing data
- Data loss impact
- Risks
- Data improvement

Benefits of

- Higher quality
- What data could be sold for
- What competitor would pay
- Expected revenue from innovative uses of data

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Valuation challenge / risk

- Data is contextual
- Temporal, dynamic
- Type of data (velocity)

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Quality

Key elements to consider:

- Technical: coding, security, redundancies, architecture
- Functional: business model, processes, completeness
- Governance: quality, accessibility, collection, storage, use, disposal, users, stewardship, security, leadership

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

low-quality

- Scrap and rework
- Inefficiencies, low productivity
- Compliance costs, fines
- Opportunity costs
- Low job, client satisfaction

high-quality

- Improved customer experience
- Higher productivity
- Reduced risks
- Increased revenue
- Competitive advantage

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Planning

- Architecture
- Modeling
- Business processes
- Budget, scope and schedule, and quality
- Trade-offs
- Planning for the long-term and short-term

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Lifecycle

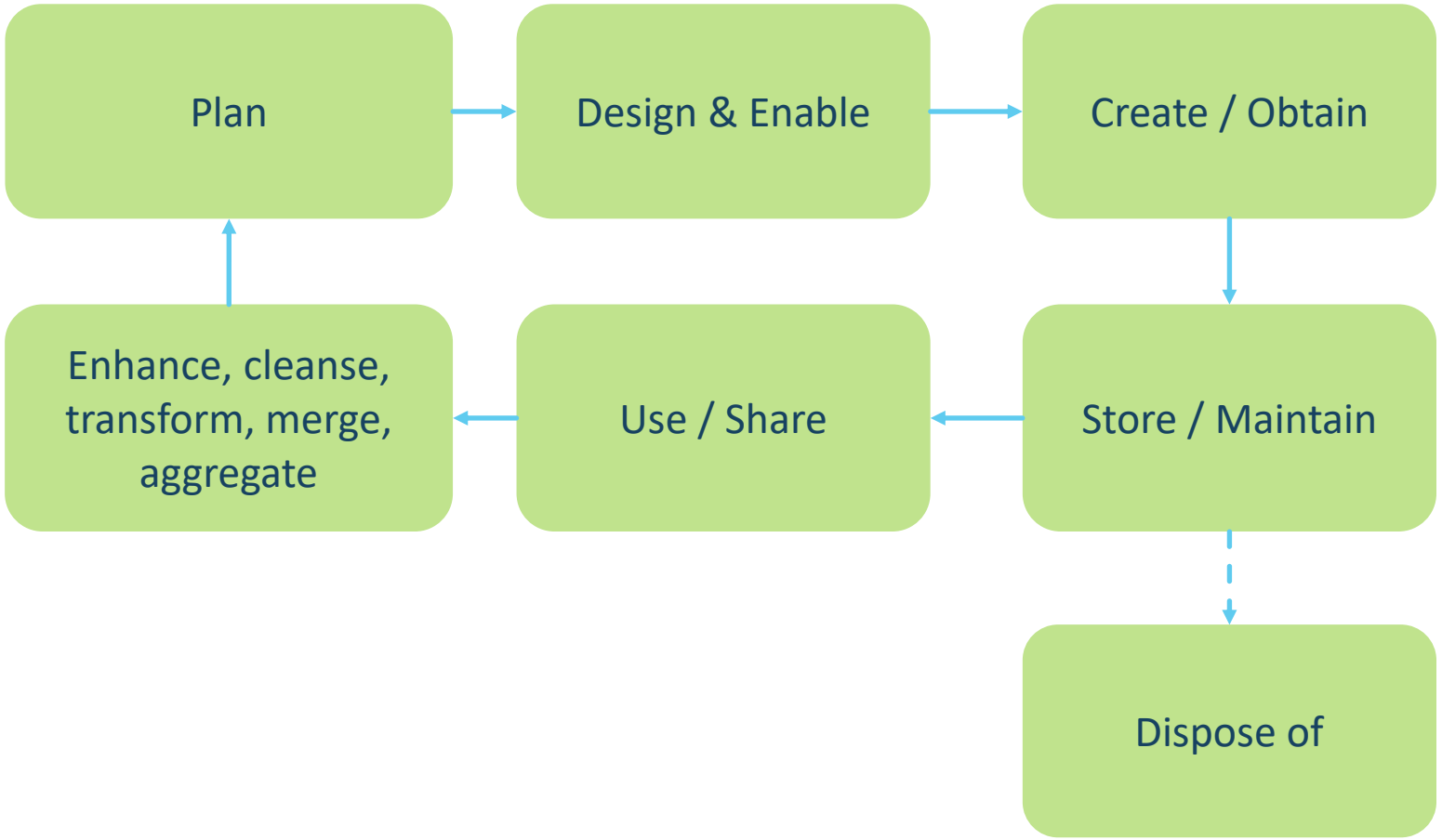
Data lifecycle is based on the product / service lifecycle.

Lifecycle includes:

- Creation and usage
- Quality management
- Metadata management
- Security management
- Emphasis on critical data

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Key activities



[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

- Regulators and legislators have increased their focus on the potential **uses and abuses** of information
  - Controls over accuracy and validity of financial data
  - Focus on data lineage and quality
  - Data privacy regulations
  - Increased awareness by the public on how data is used

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Management – Leadership and commitment [1] [2]

THE LEADER'S DATA MANIFESTO

YOUR ORGANIZATION'S BEST OPPORTUNITIES FOR **ORGANIC GROWTH** LIE IN **DATA**

DATA OFFERS ENORMOUS UNTAPPED POTENTIAL TO CREATE COMPETITIVE ADVANTAGE, NEW WEALTH AND JOBS; IMPROVE HEALTH CARE; KEEP US ALL SAFER; AND OTHERWISE **IMPROVE THE HUMAN CONDITION**

ORGANIZATIONS ARE FAR FROM BEING DATA-DRIVEN. MOST COMPANIES:

1. DON'T FULLY KNOW WHAT DATA THEY HAVE OR EVEN WHAT DATA IS MOST IMPORTANT
2. CONFUSE "DATA" WITH "INFORMATION TECHNOLOGY" OR DIGITALIZATION, LEADING THEM TO MISMANAGE BOTH
3. LACK ANY SORT OF DATA VISION OR STRATEGY DEFINING HOW DATA CONTRIBUTES TO THEIR BUSINESS
4. UNDERESTIMATE THE EFFORT REQUIRED TO MANAGE DATA AND LACK THE ORGANIZATIONAL STRUCTURE TO DO SO

MANY COMPANIES ARE SUCCEEDING WITH SMALL-SCALE ANALYTICS, GOVERNANCE, QUALITY AND OTHER EFFORTS.

STILL, WE FIND **NO EXAMPLES OF FUNDAMENTAL, LASTING, COMPANY-WIDE CHANGE WITHOUT COMMITTED LEADERSHIP** AND THE INVOLVEMENT OF EVERYONE AT ALL LEVELS OF THE ORGANIZATION.

... AND WE ARE FULLY AWARE HOW DIFFICULT IT WILL BE TO UNLOCK DATA'S POTENTIAL, UP AND DOWN THE ORGANIZATION CHART.

THEREFORE, WE URGE EVERYONE TO LEAD CHANGE

BOARDS, SENIOR EXECUTIVES AND SENIOR LEADERS: CHALLENGE YOUR PRECONCEIVED NOTIONS OF DATA.

See data not as the details buried in the bowels of IT and in your computer systems, but as a **source of unlimited, new opportunity**. Realize that data's potential isn't just for specialists, such as data scientists, but for everyone and a way for your company to **truly distinguish itself** from competitors. See data as a way for you personally to **leave an enduring legacy**. Consider what it would mean to put data assets on your balance sheet.

Leadership owes their shareholders and constituents a data vision. As a first step, focus on these areas:

1. Take better care of data, with a focus on the quality of your most important data. **2.** Try out the many ways to put your data to work and gain a competitive advantage. **3.** Advance a management system better suited to the rigors of data.

EVERYONE WHO NEEDS DATA TO DO HIS/HER JOB: BECOME DATA PROVOCATEURS TO DRIVE CHANGE.

Opportunities abound, so choose an area or two that interests you. It could be improving data quality; discovering a deeper analytic method; developing a new metric; delivering an idea that quantifies the hard-dollar value of data; or using data to build bridges with other departments.

DATA PROFESSIONALS:

Be more proactive, communicate with your business counterparts, sell the concepts, become a data mentor, and help people create their own success stories.

GET ON WITH THE WORK

These are exciting and perilous times: exciting because data offers opportunities to create competitive advantage, enhance existing products and services (and create new ones), better understand customers and reduce costs. And perilous because fixing what's broken will be arduous, and those who wait too long may find themselves severely disadvantaged.

Give this manifesto deep consideration.

SHARE IT. DEBATE IT.

And make it work in your organization.

www.dataleaders.org

DR. NINA EVANS · JOHN LADLEY · DANETTE MCGILVRAY · KELLE O'NEAL · JAMES PRICE · TOM REDMAN

© 2017 dataleaders.org

[2] https://dataleaders.org/manifesto/

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Risk

Data should be available, relevant, accurate, complete, consistent, timely, usable, meaningful, understood

Risks

- Low quality data (inaccurate, incomplete, out-of-date)
- Misunderstood or misused data
- Information gaps

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Management Strategy

Deliverables/Components

- Data Management Charter
- Data Management Scope Statement
- Data Management Implementation Roadmap
- Data Governance policy, program and procedures

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Data Management Framework

- A functional DM framework help a data-driven organization to apply DM concepts, use a common vocabulary and serve as the fundamental reference guide for the Certified Data Management Professional

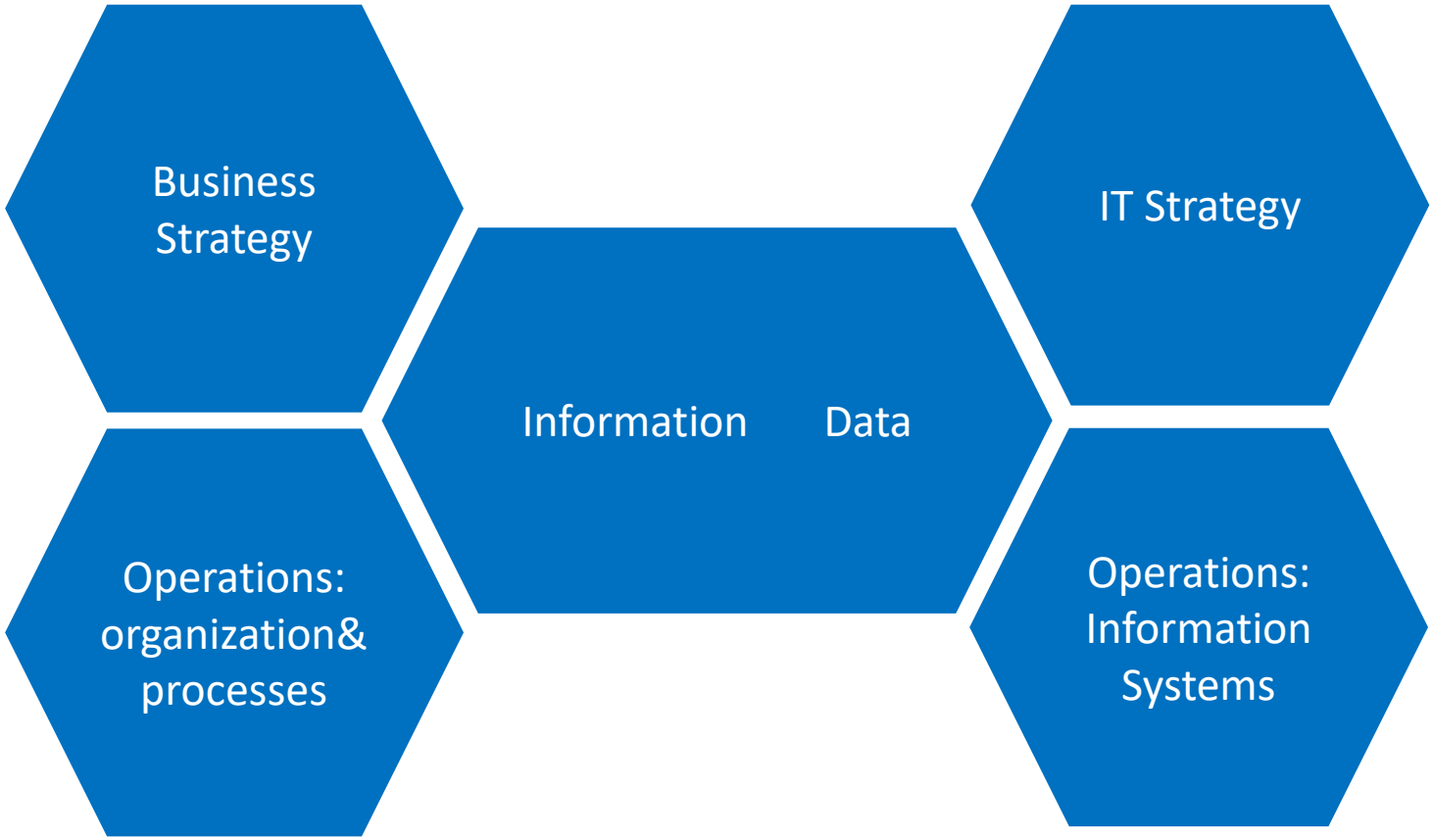
[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Models

1. Strategic Alignment Model
2. Amsterdam Information Model
3. **DAMA DMBOK framework – DAMA Wheel Model**
4. DMBOK Pyramid (Aiken)
5. DAMA Data Management framework evolved
6. DAMA Data Management Function framework
7. DAMA Wheel Model evolved

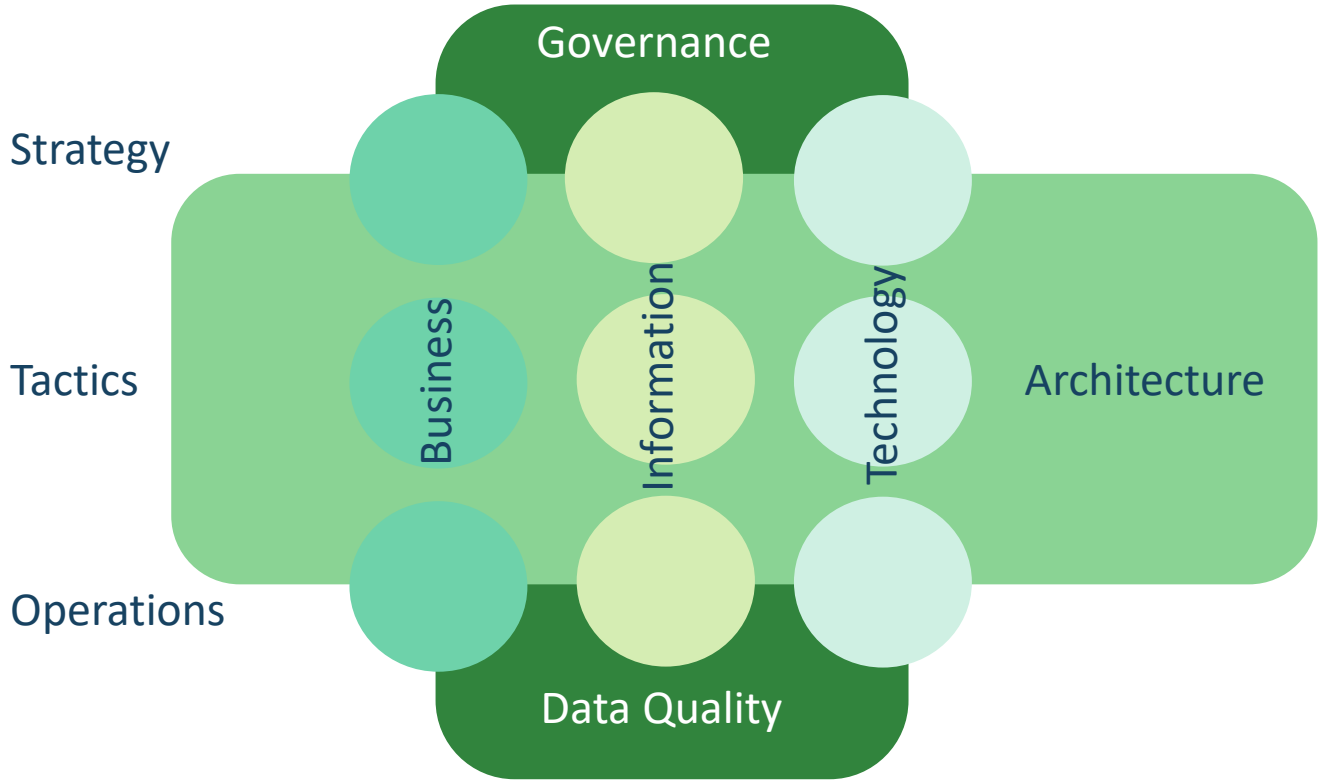
[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

1. **Strategic Alignment Model** (Henderson-Venkatraman, 1999)



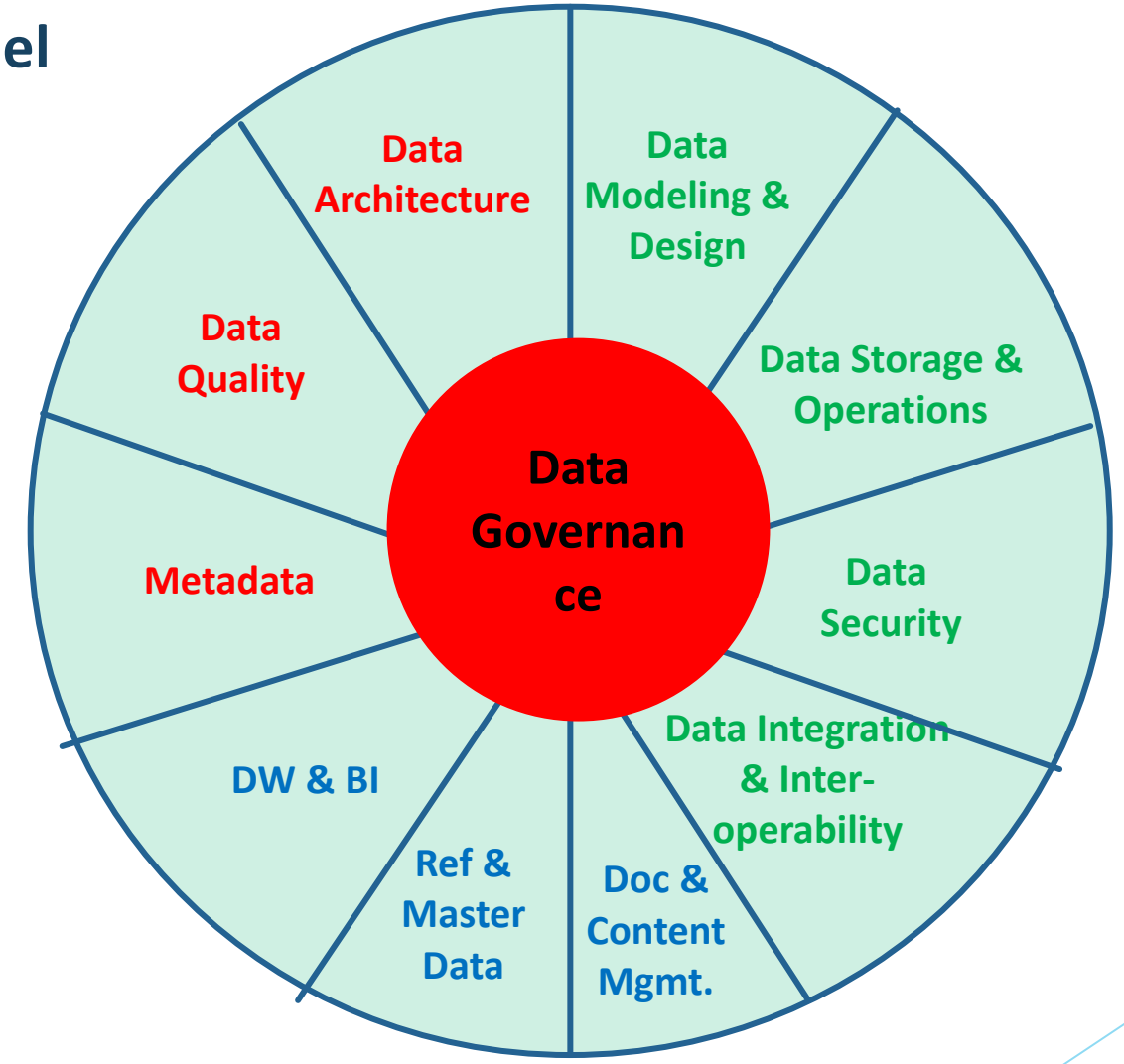
[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

2. The Amsterdam Information Model



[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

3. The DAMA Wheel



[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

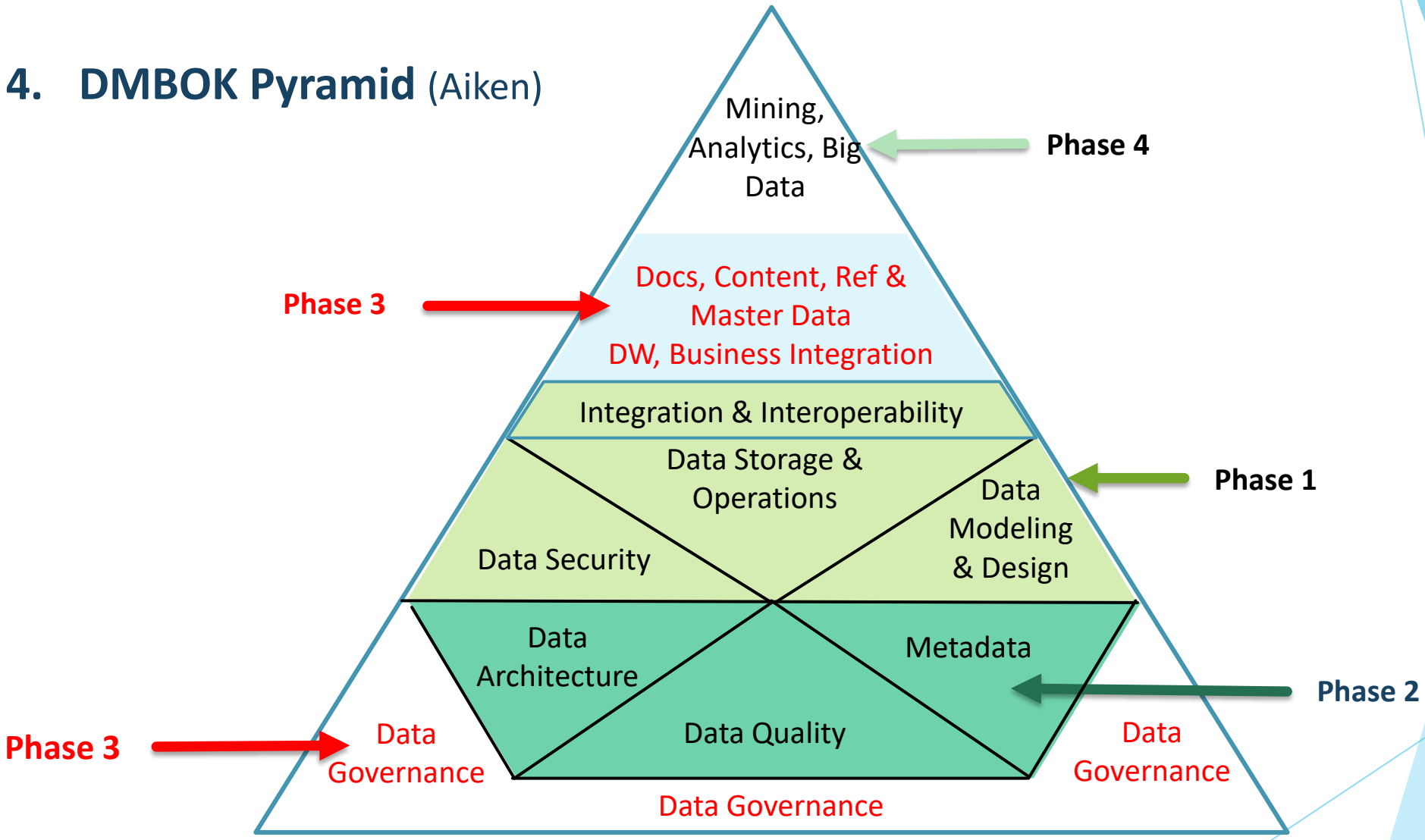
4. **DMBOK Pyramid** (Aiken)

States where organizations have reliable data and processes to support strategic business goals. Progression steps:

- deploy an application with **database** capabilities
- implement a reliable **Metadata**
- implement a consistent **DA** for integration, interoperability
- implement **DG** practice to support for Data Management activities
- deploy **Data Analytic** capabilities

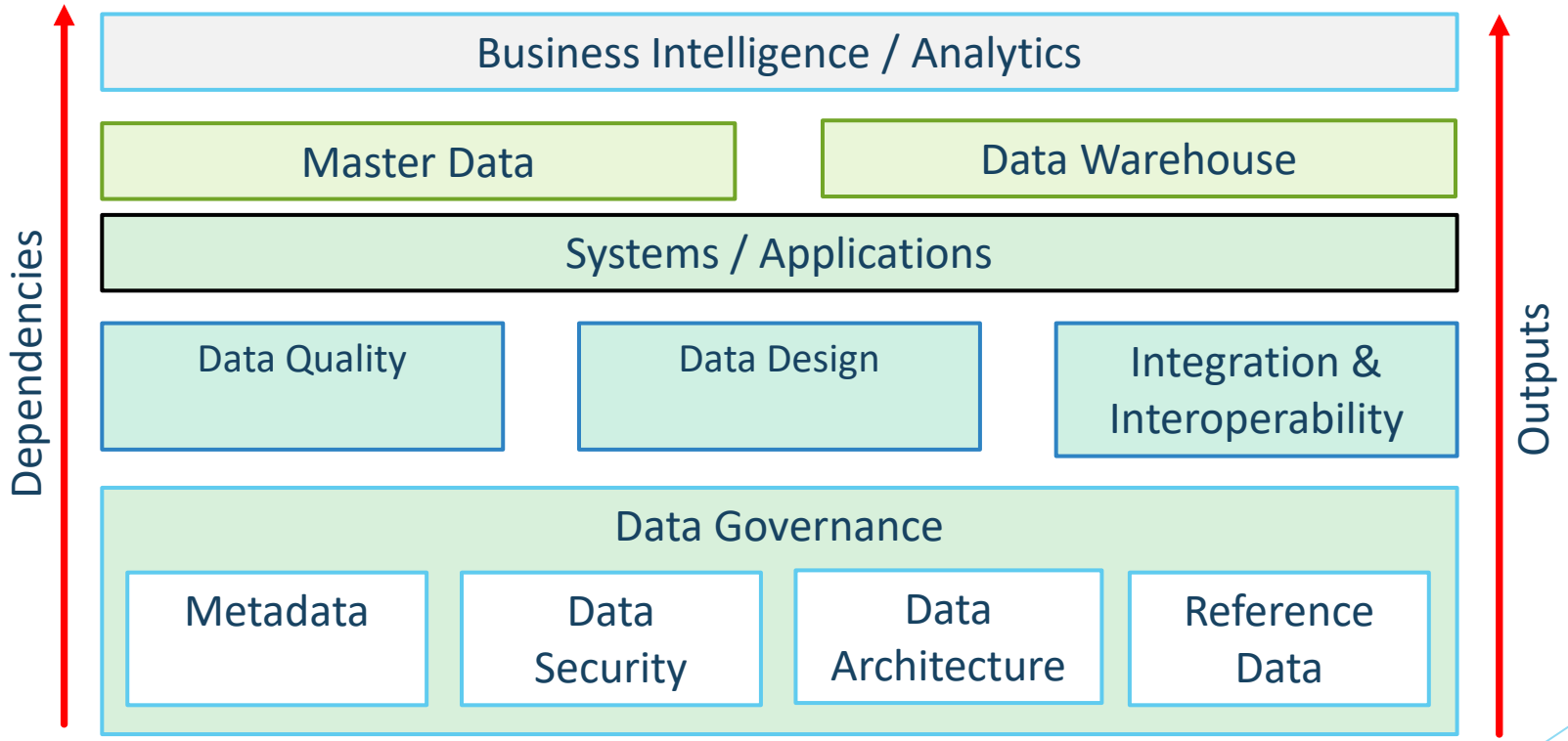
[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

4. DMBOK Pyramid (Aiken)



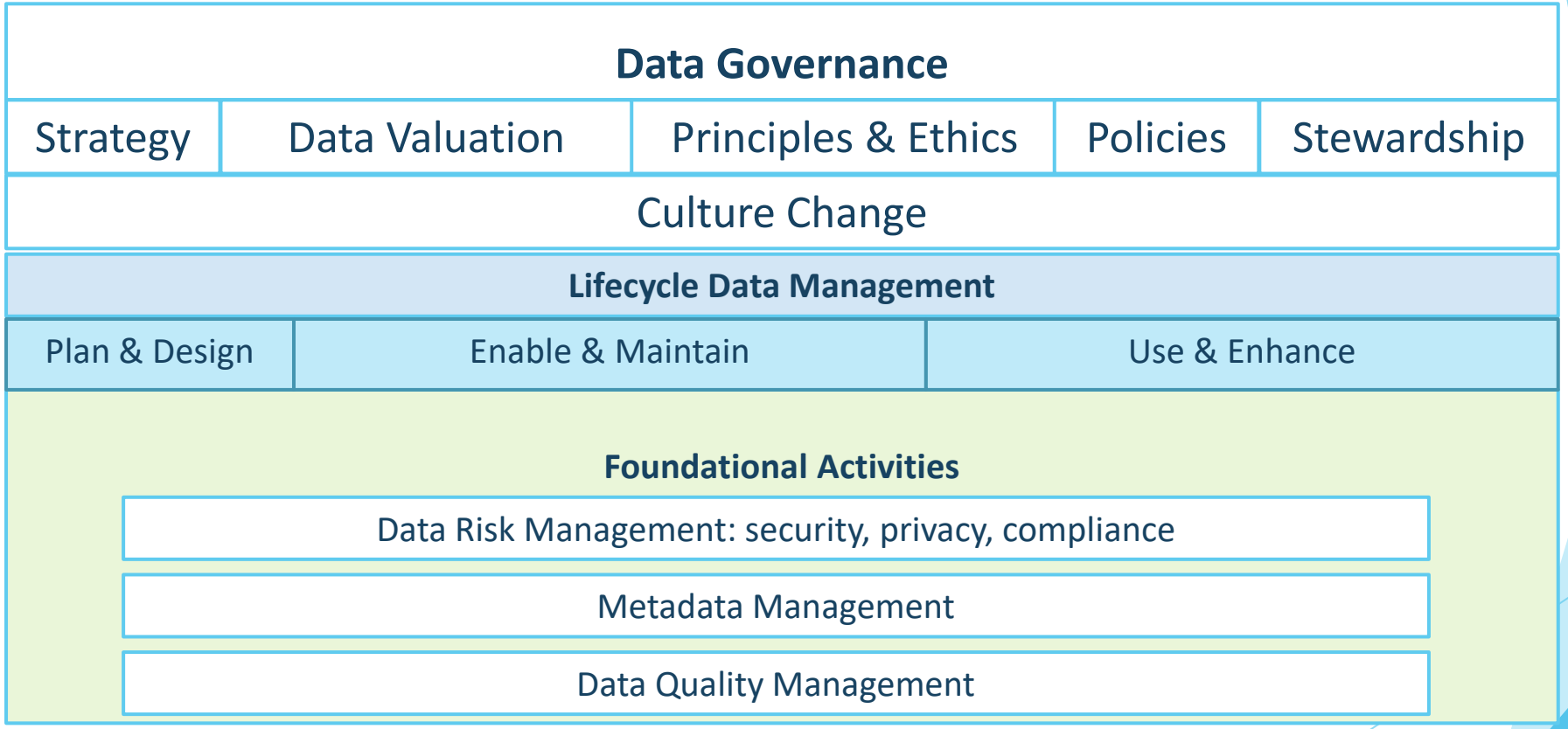
[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

5. DAMA Data Management Framework Evolved[1] (S.Geuens)

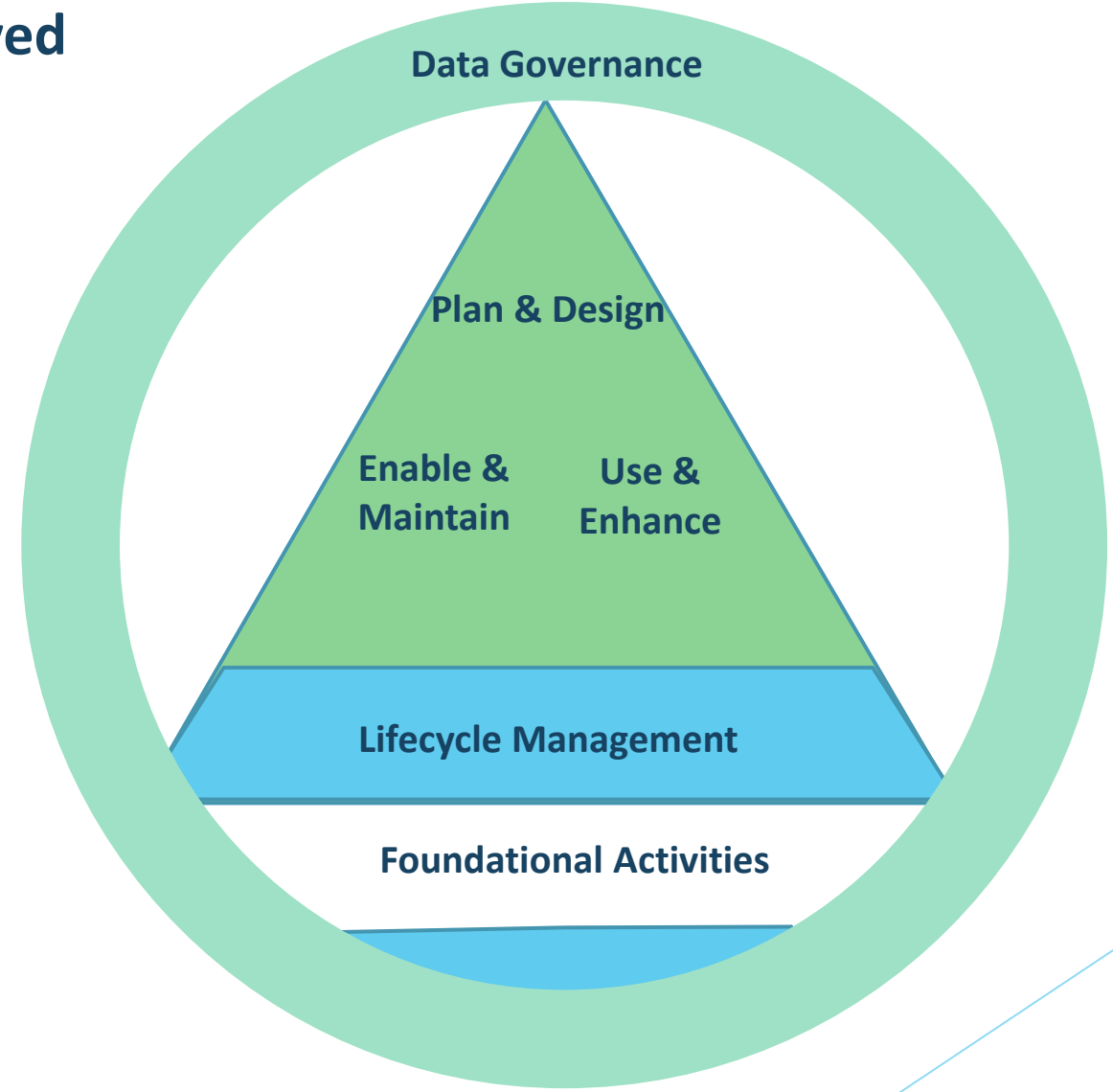


[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

6. DAMA Data Management Function Framework



7. DAMA Wheel evolved

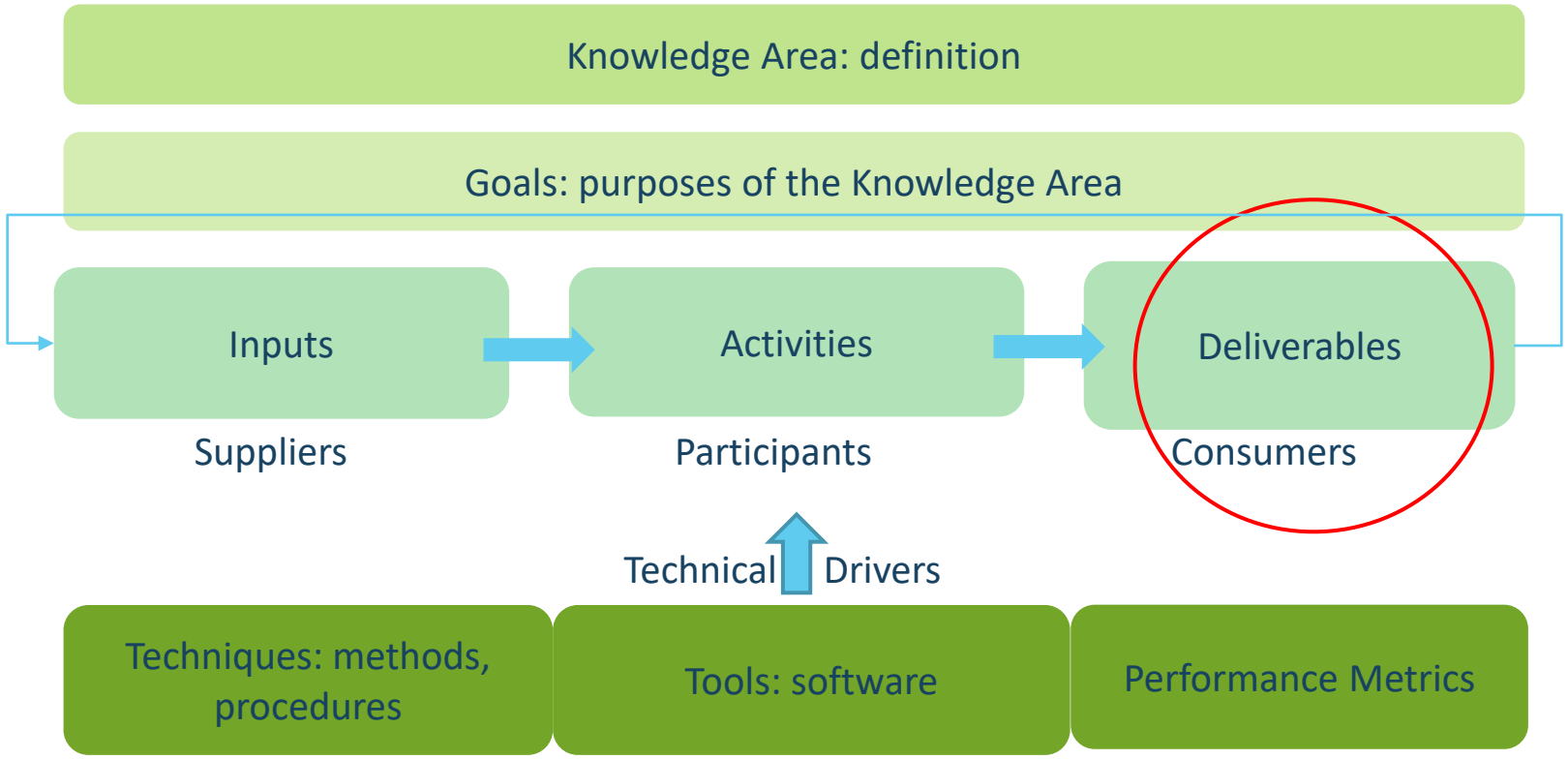


[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Environmental Factors Hexagon

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Knowledge Area Context Diagram (generic)



[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

DAMA-DMBOK - Knowledge Areas

1. Data Governance
2. Data Architecture
3. Data Modeling and Design
4. Data Storage and Operations
5. Data Security
6. Data Integration and Interoperability
7. Document and Content Management
8. Reference and Master Data
9. *Data Warehousing and Business Intelligence*
10. Metadata
11. Data Quality

***Note**: for additional topics and/or bibliography, and/or works cited, refer to the DAMA-DMBOK2*

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA