Architected Data Governance

Guy Tozer, Doriq Associates

DG Conference Europe 2009

Introduction

- Background
- Speaker Introduction
- Audience Profile
- Purpose and Focus of the Presentation
- Ground-rules and Capabilities

Background

- Enterprise Architecture and DG are closely linked as concepts, but are rarely, if ever, managed in a cohesive way
- The intersection of these two disciplines, Data Governance Architecture, is often poorly understood leading to:
  - EA remaining a theoretical exercise (at least as far as data is concerned)
  - Wasted effort within the DG community, due to lack of enterprise context

Structure of the Presentation

- Introduction
- Part I: The Problem Domain
- Part II: EIA in Context
- Part III: EIA and DQM –The problem in detail
- Break
- Part IV: Bringing EIA and DQ together
- Part V: Tools and Techniques
- Wrap-up

Speaker Introduction

- Guy V Tozer
  - Author – books on:
    - Metadata Management
    - Information Quality Management
  - Consultant
  - Lecturer
  - Entrepreneur
  - Many years experience of both DQM, Data Governance and Enterprise Architecture

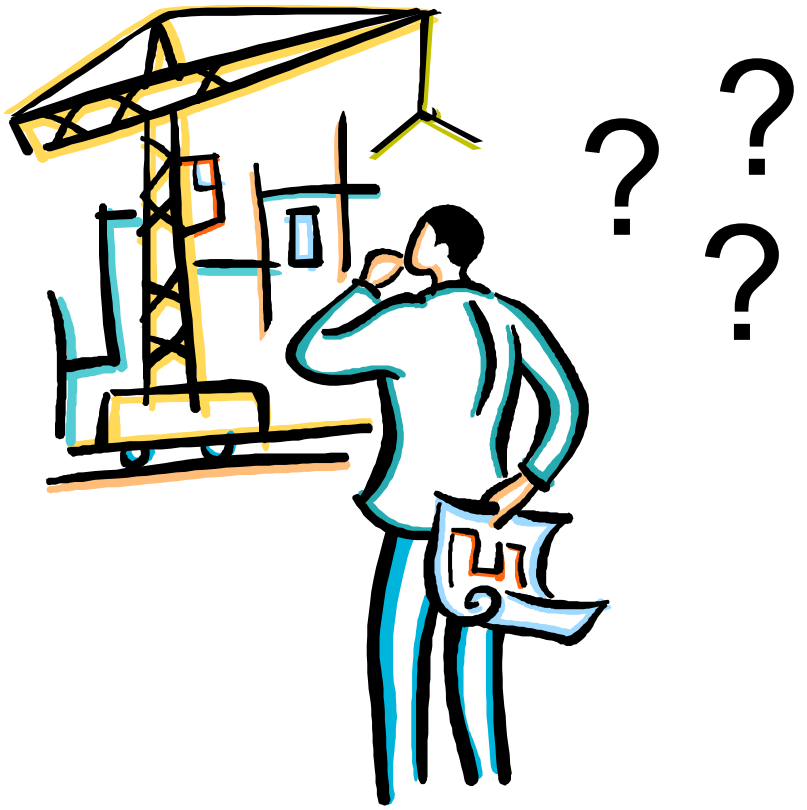
Audience Profile

- Some expertise in Data Governance practices
- Concerned about long-term rigour of DG processes
- Data and metadata-literate
- Seeking ways to improve the **robustness**, **flexibility**, **efficiency** and **effectiveness** of their DG practice

Content of the Presentation

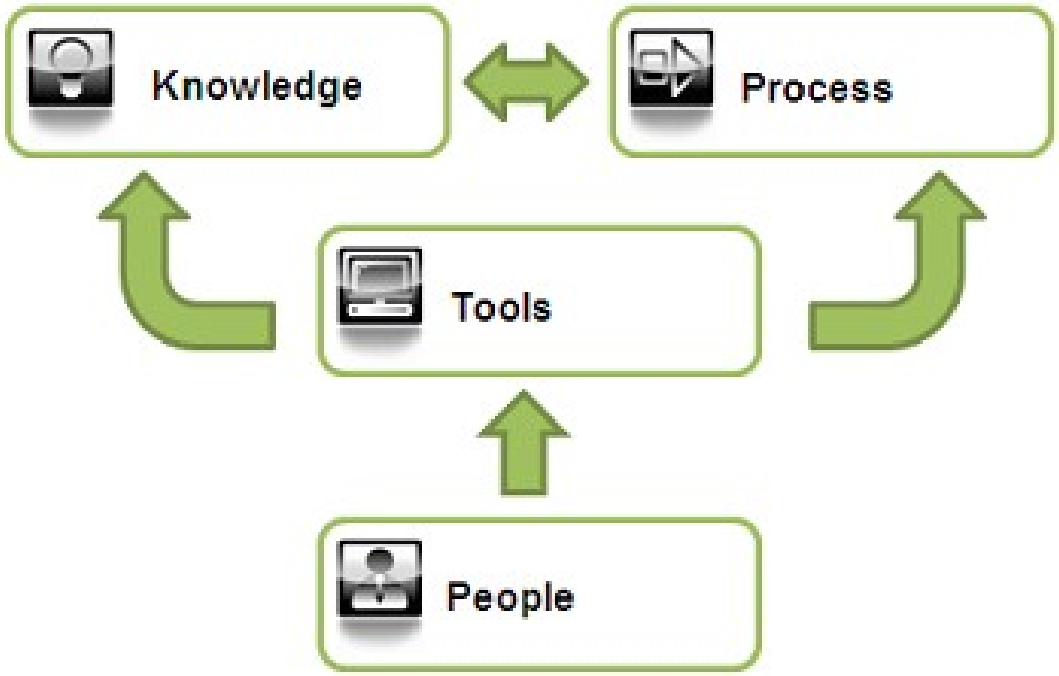
- What do we mean by an Architected Approach?
- How does it help us to address the challenges of Data Governance?
- What knowledge and processes support Architected DG?
- How do we go about setting up and sustaining such an approach?

Part 1: What is Architecture?



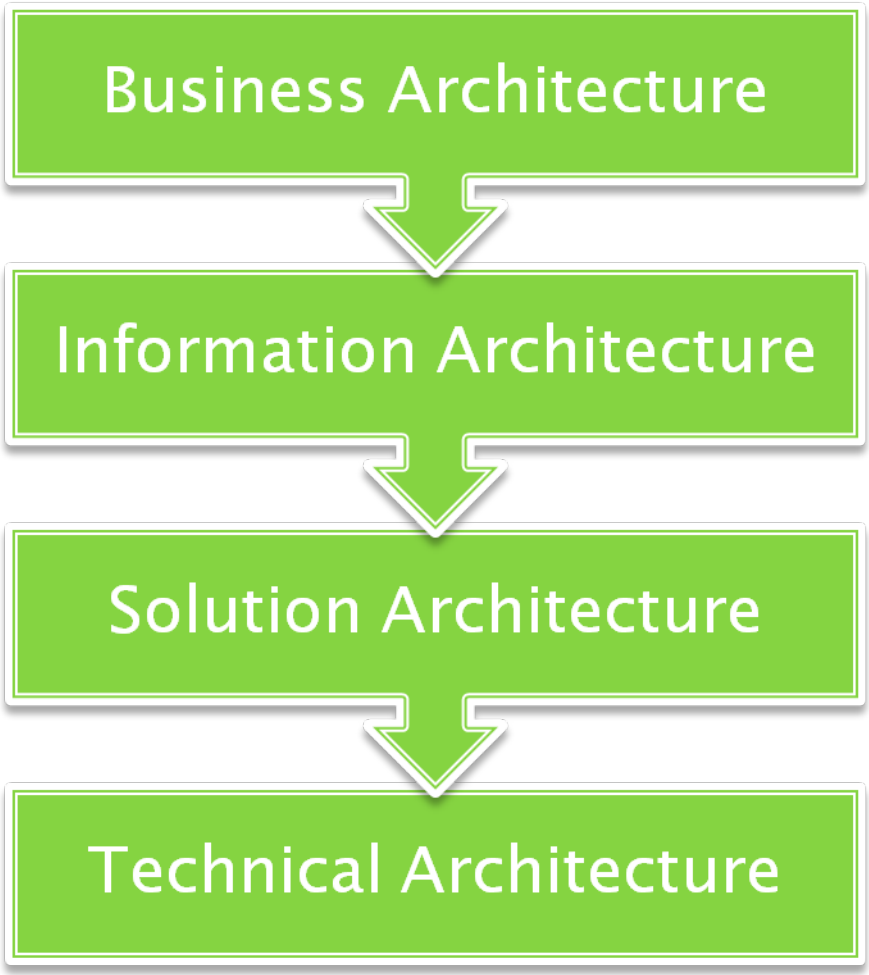
doriq

Components of 'Good' Architecture



Architecture Layers Overview

- Although many different approaches to EA exist, a common thread running through them is the need for formal connection between the so-called architectural layers

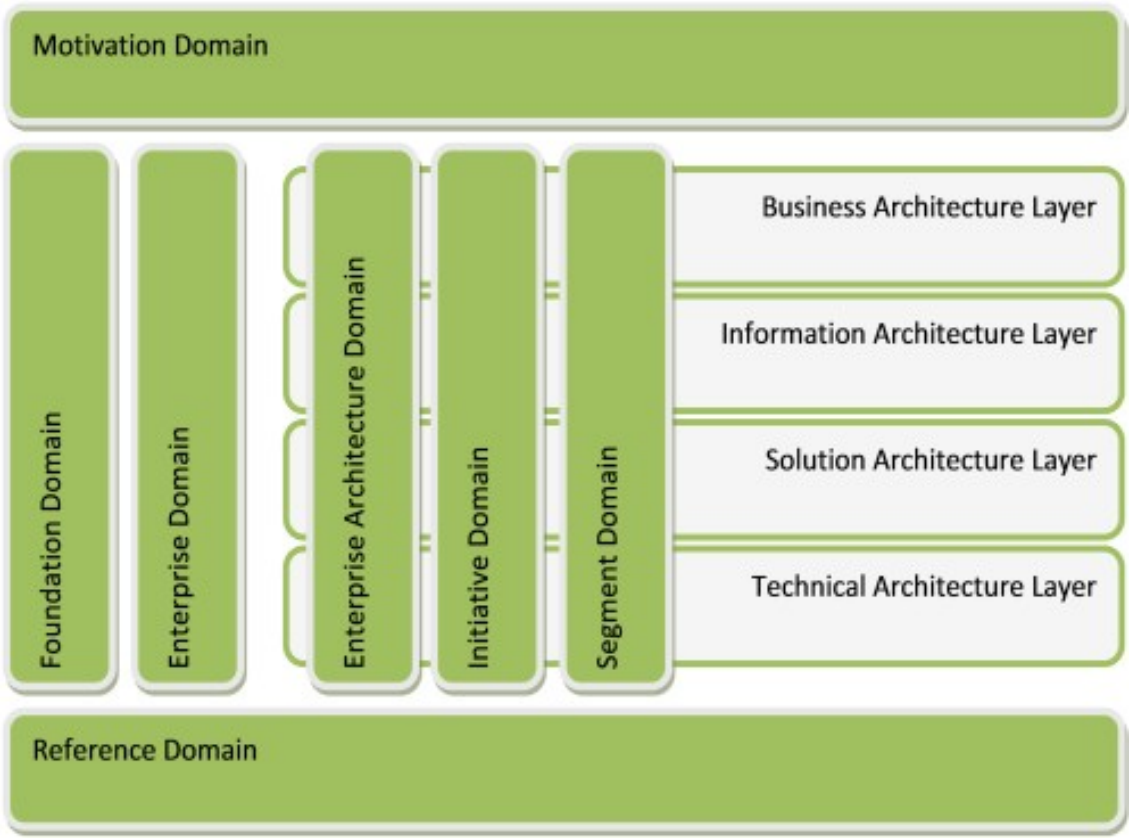


Scope of the EA Layers

- Business Architecture, defining the processes, operation, context and motivation of the business;
- Information/Data Architecture, prescribing and defining the types of information relevant to the business;
- Systems Architecture, describing the logical behavior of systems/applications in the context of the business;
- Technical Architecture, describing the physical infrastructure required to provide and support these systems.

Knowledge Domains

- Architectural Knowledge may be split into domains for ease of understanding:



The Zachman Framework

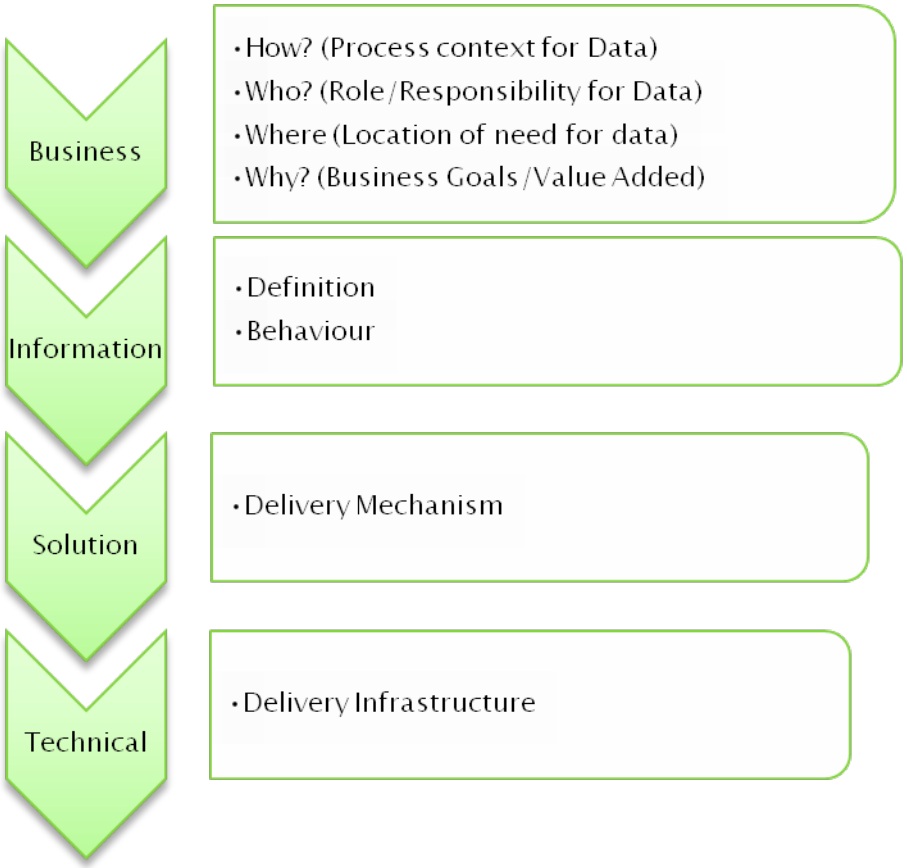
- The Zachman Framework was an early (and still authoritative) way of organizing EA knowledge.
- John Zachman proposed an (initially) two dimensional grid:
  - the rows represented increasing levels of precision, from abstract (a rough textual description of something) to concrete (the real-world existence of something)
  - the columns represented various categories of knowledge, represented by interrogative words

The Zachman Interrogatives

- Why? (MOTIVATION)
- What? (DATA)
- How? (FUNCTION)
- Where? (NETWORK –incl. Location)
- Who? (PEOPLE)
- When? (TIME)

Acknowledgement: John Zachman, ZIFA

Data Governance and the Interrogatives



NB: Each layer provides context for those below

Architectures of Architecture

- Town Planning analogy
- DG Architecture overlaps with following domains:
  - Data Management
  - Business Process Management
  - Security
  - Organisational Design
  - Technical Data Architecture (e.g. DBMS, Data Integration, etc.)

Part 2: How does Architecture help?

Architectural Drivers

- An Architected Approach (to any complex problem):
  - Provides rigour and authority
  - Improves flexibility and responsiveness to change
    - Understanding impacts
  - Facilitates ‘encapsulation’
  - Sustains and support reusability
  - Lowers aggregate costs

Data Governance Focus Areas

- Policy, Standards, Strategy
- Data Quality
- Privacy, Compliance, Security
- Architecture/Integration
- Data Warehouses and BI
- Management Alignment

Motivational Complexity

- Each Focus Area has a set of motives, adding perceived value through DG activities
- Motives (and other components) for each Focus Area can interrelate in several ways:
  - Pre-requisite
  - Composition
  - Shared Component
  - Conflicting Component

The Data Governance Journey

- Find good people to ‘govern’
- Understand your current situation
- Develop a strategy for Data Governance
- Determine the value of your data
- Understand risk
- Measure ongoing effectiveness

Adapted from “Six Steps to Data Governance”, Steven Adler

Data Governance People

- Must provide strong leadership
  - Represent the enterprise
  - Need authoritative enterprise viewpoint to do this
- Roles and Responsibilities need to be carefully designed
  - Again, enterprise perspective vital for this

Question: Where does this enterprise viewpoint come from?

The AS-IS Situation

- Need to understand overall maturity levels for Enterprise Data Governance
- Need to have a clear, unambiguous picture of existing practices
- Need to understand the problems caused by existing practices
- Used as basis for developing DG Roadmap
- DG Roadmap used as critical input to the DG Strategy

Data Governance Strategy

- A DG Strategy must address the fundamental issues:
  - Lack of Enterprise-level DG Organisation Structures
  - Lack of full life-cycle perspective on requirements
  - Lack of full life-cycle perspective on risk
  - Data Quality assessment not based on formal metadata
  - Poor technological support for DG
  - Short-term approach to governance model

The Value of Data

- Data Governance fulfils the primary function of preserving and enhancing the VALUE of data across the enterprise
- Understanding what constitutes Data Value is therefore of primary importance to successful DG
- Commercial value is determined by ‘how much people are willing to pay’
- In an internal ‘market’ we need a more subtle approach

So what is Data Value?

- Data Value is often expressed in nebulous terms:
  - “The data is under control”
  - “We have high quality data”
  - “The data is fit for purpose”
  - “We trust the data”

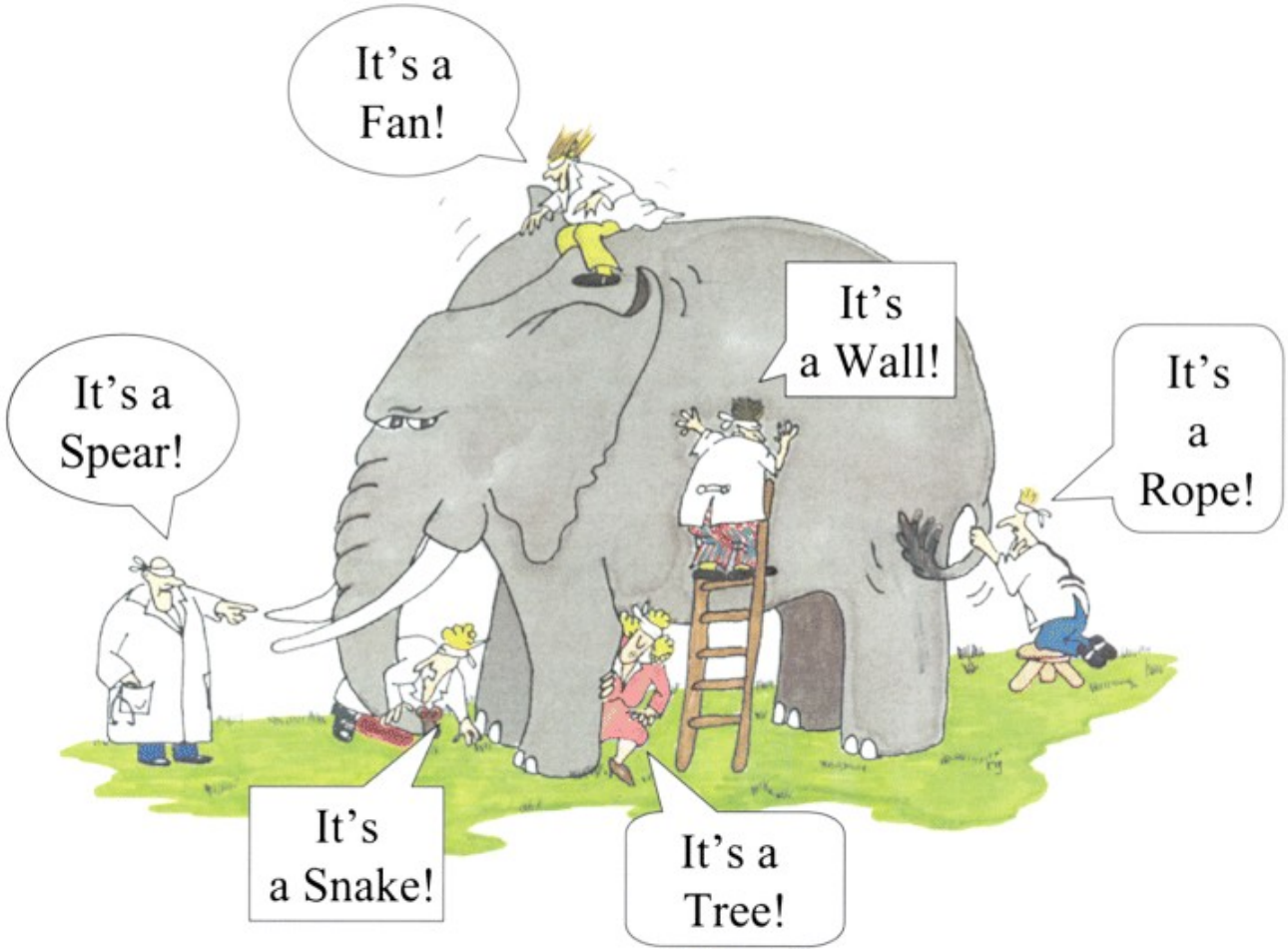
A critical eye on Data Value

- These statements lead us to a number of challenging questions:
  - Under whose control? On behalf of the enterprise?
  - What exactly is meant by ‘high quality data’?
  - How can we state that data is fit for purpose without a profound understanding of ‘purpose’?
  - Is trust in the data universal, or selective?

Understanding Risk

- As with quality, risk must be assessed with an enterprise perspective.
- Historical problems with data must be gathered, analysed and assessed for their enterprise significance
- Risk and value are conjoined and interdependent, and must therefore be managed at the same level (the enterprise level)

Non-enterprise risk assessment



doriq

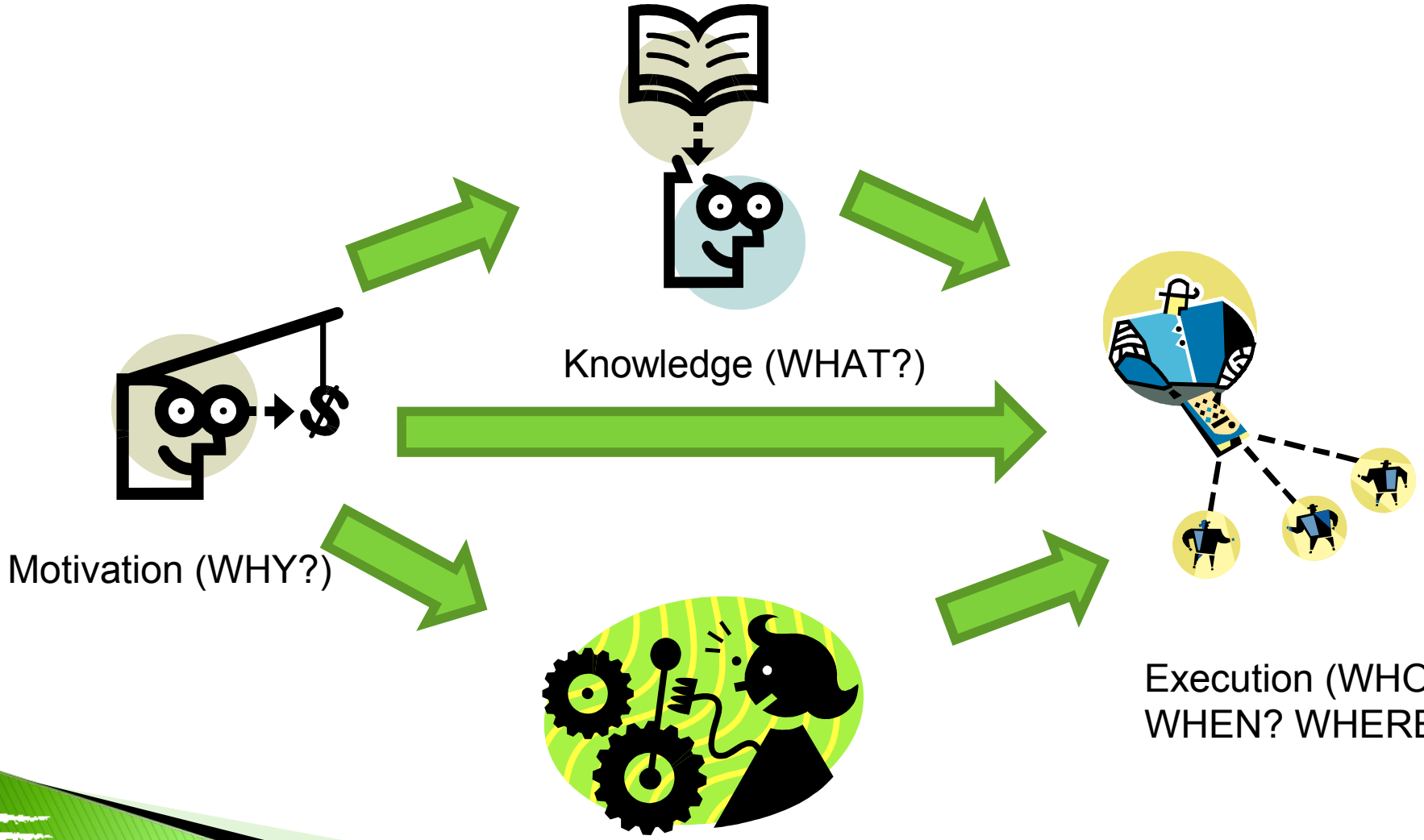
Measuring Ongoing Effectiveness

- It is important to measure how well the data fulfils the need expressed in the business MOTIVATION
  - Focus on real, enterprise Data Value
- This should not be confused with:
  - How well the PROCESS of DG is performed;
  - The amount of DATA under the control of DG
  - The basic accuracy and correctness of the data content

Summary: Architected DG

- Let us re-examine the Zachman Interrogatives:
  - Why?
  - What?
  - How?
  - Who?
  - Where?
  - When?
- For DG (and other activities) there is a chain of interdependencies to consider.

The Architectural 'Chain' for DG



doriq

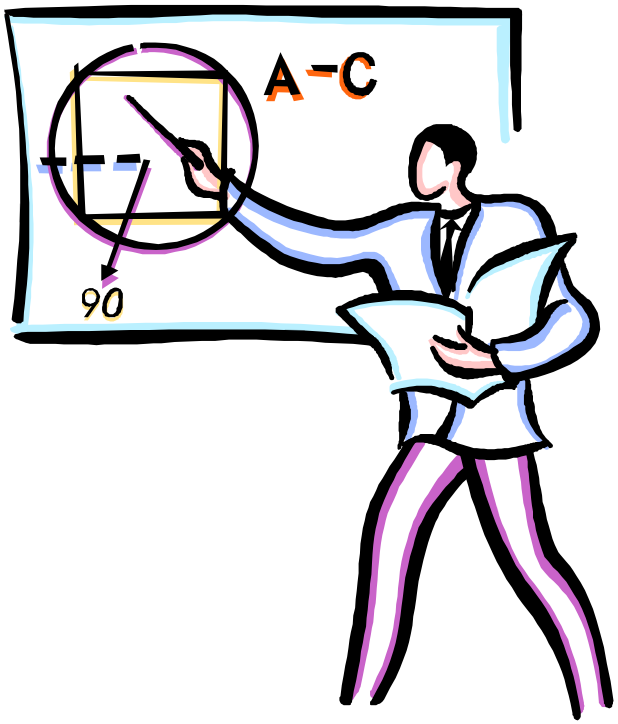
Chain Interdependencies

- A proper understanding of the Motivation is imperative to:
  - Understand the knowledge required
  - Design an effective process for DG
  - Execute the process and manage the knowledge, at the right time, with the right resources, in the right place

Questions?

Break

Part 3: Metamodels and Processes



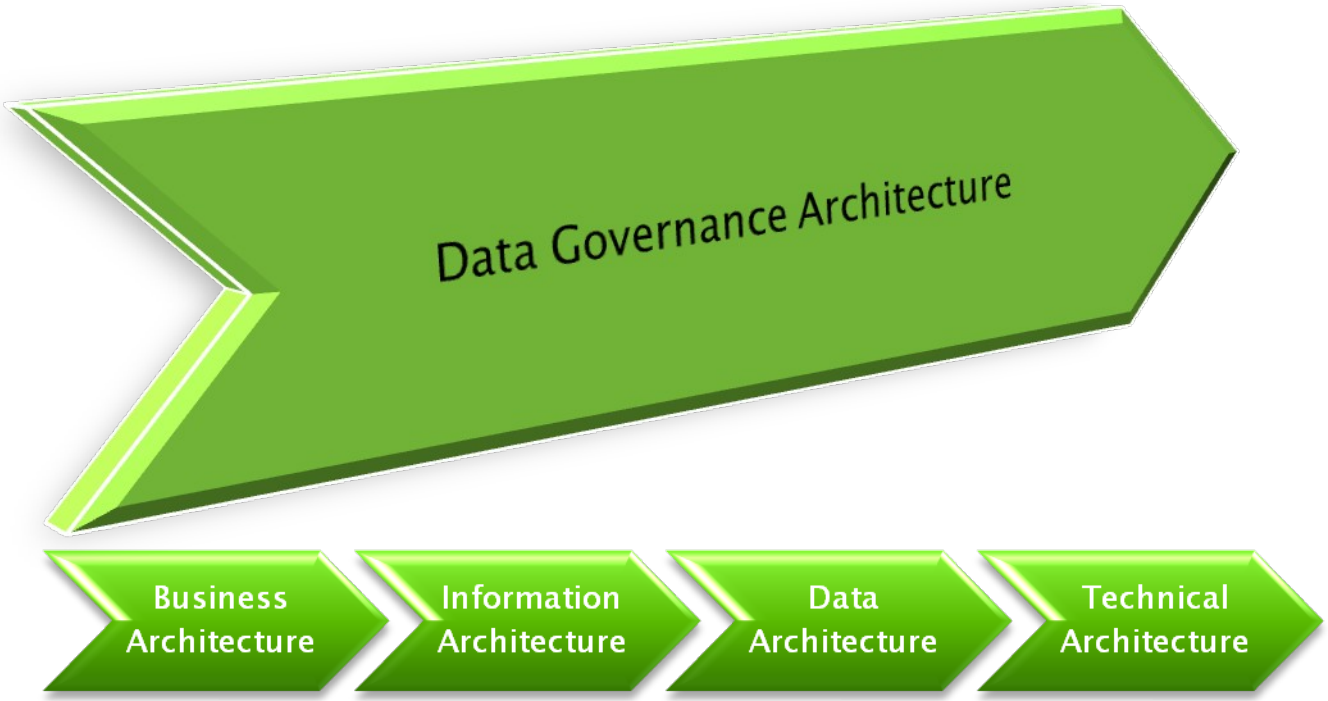
doriq

Metamodels for Data Governance

- There are a number of Knowledge Models or ‘metamodels’ which can be used to formalise the various concepts needed to perform Data Governance functions
- The DG Domain is split into a number of inter-related sub-domain, each of which is described by a metamodel

The Data Governance Domain

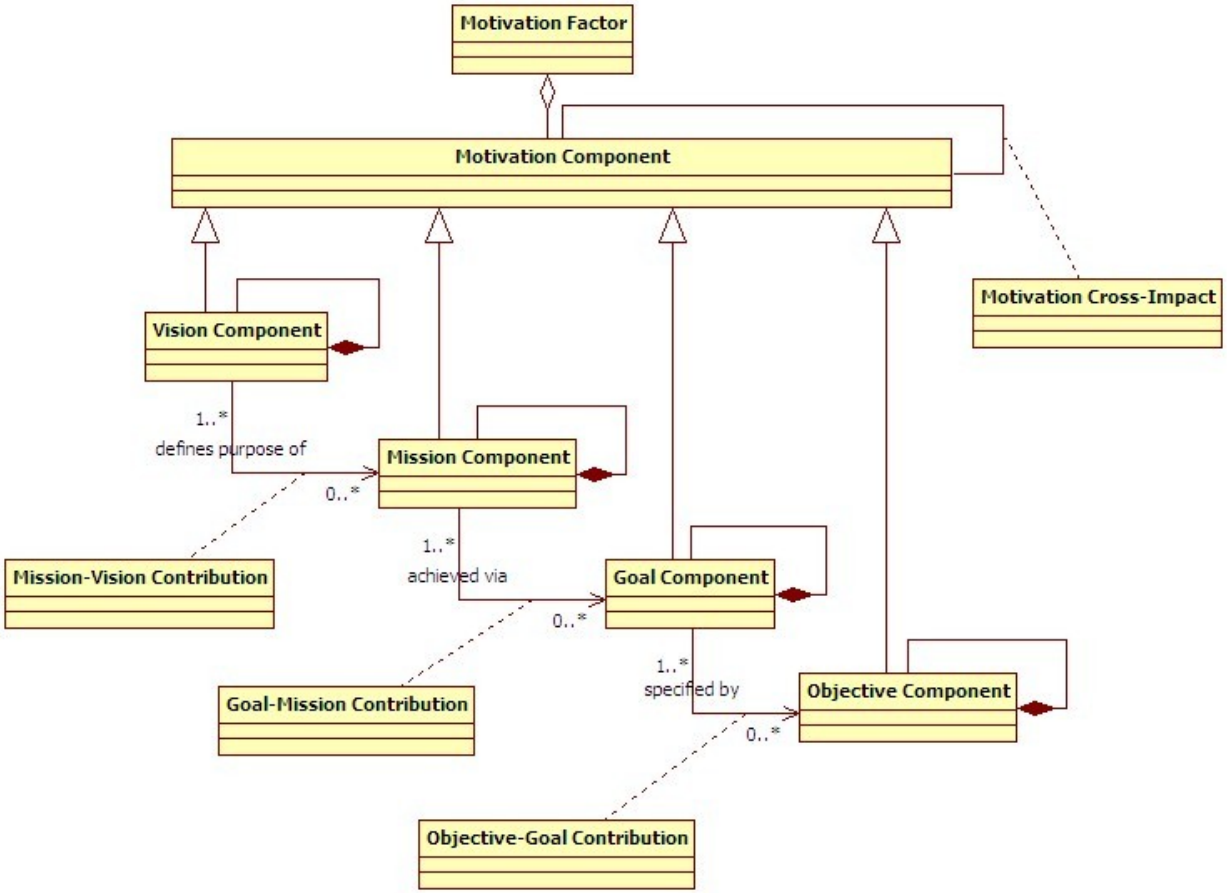
- The DG architectural domain has close links with other architectural domains:



Data Governance Sub-Domains

- Motivation Sub-Domain
- Business Strategy Sub-Domain
- Business Architecture Sub-Domain
- Information Architecture Sub-Domain
- Business Capability Sub-Domain
- Data Value Sub-Domain
- Process Sub-Domain
- Measurement Sub-Domain

Motivation Sub-Domain

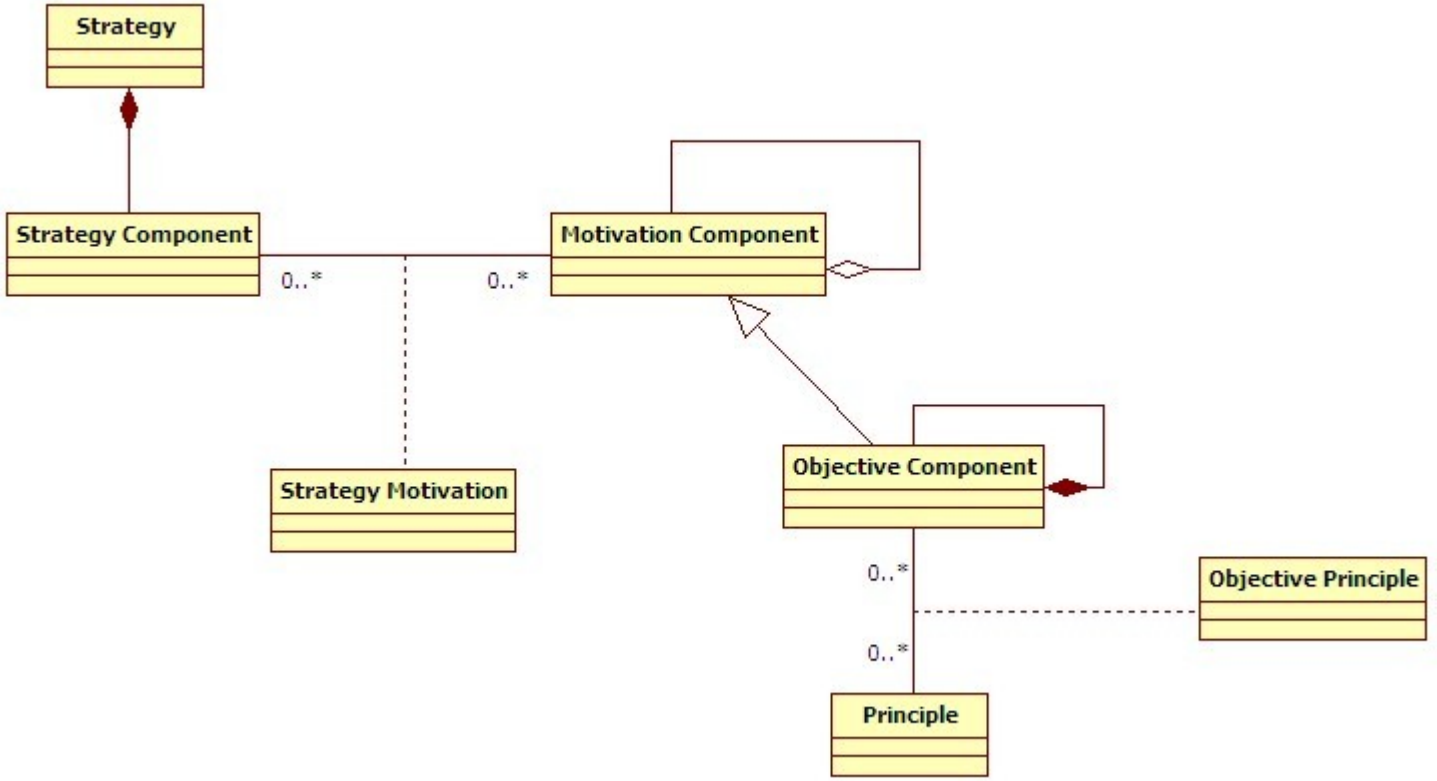


doriq

Motivation Context

- Provides standardised way of analysing and recording enterprise:
  - Vision
  - Mission
  - Goals
  - Objectives
- Provides mechanism for recording the links between them
- Ultimately leads to rigorous understanding of real data value (contribution to corporate goals)

Business Strategy Sub-Domain

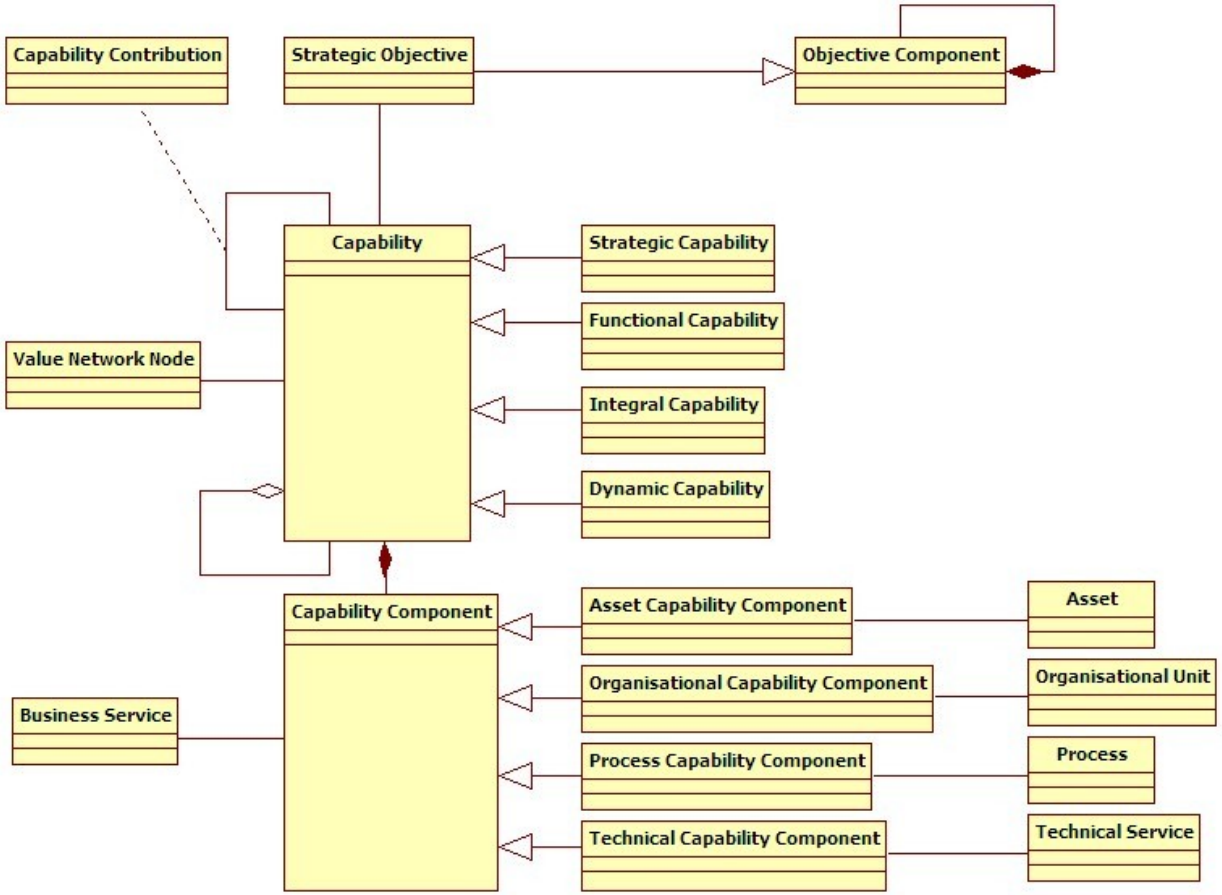


doriq

Business Strategy Context

- Parts of the strategy are driven by the enterprise objectives
- Links the WHAT with the WHY
- Standardises the link between strategy and principles

Business Capability Sub-Domain

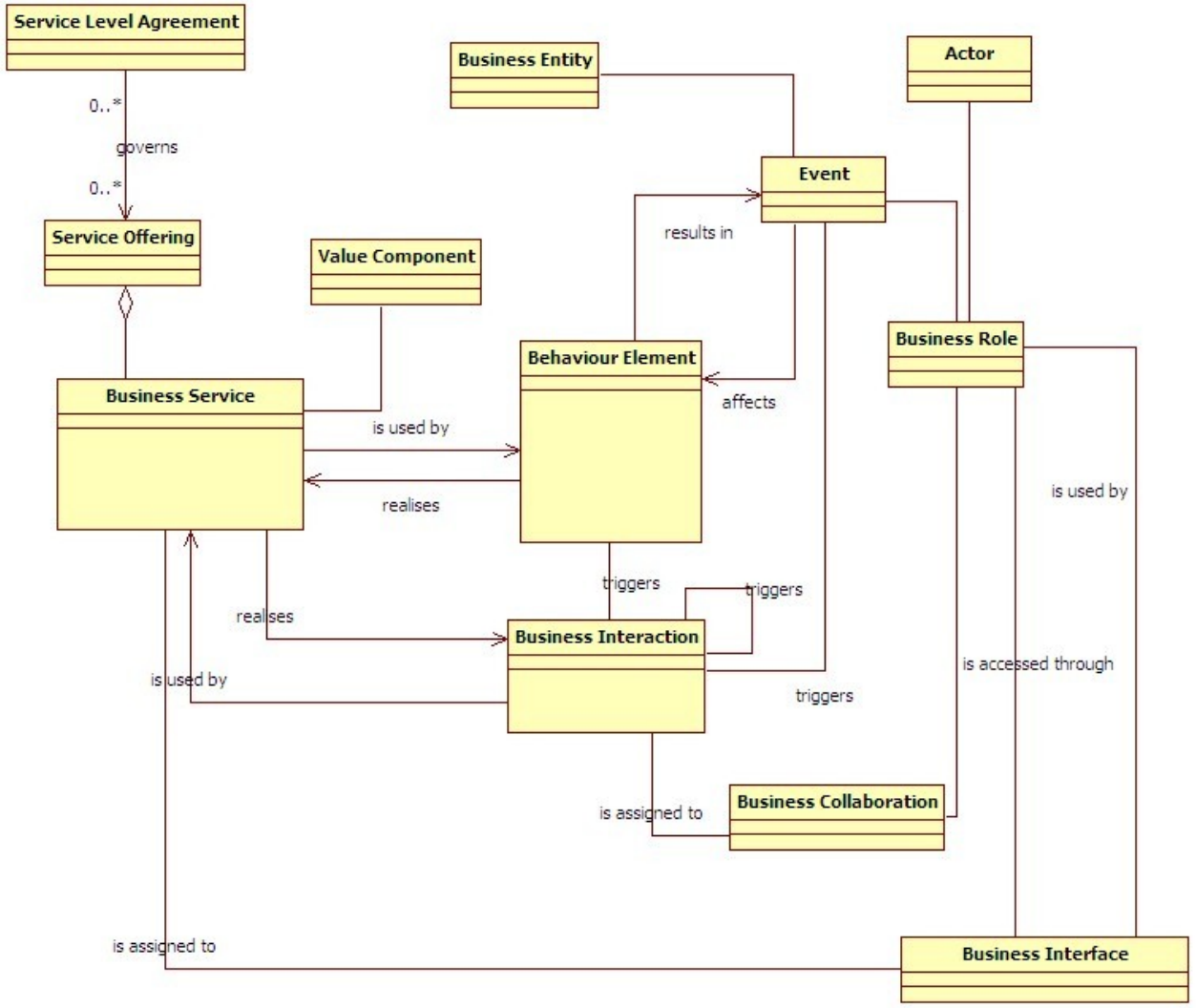


doriq

Business Capability Context

- Takes strategic objectives and links them to the business capabilities
- In DG context, capabilities such as ‘effective DG’ will provide value, through use of components:
  - Asset-based (Data)
  - Organisation-based (e.g. Data Stewardship)
  - Process-based (e.g. Controlled entity life-cycle)
  - Technical Service-based (e.g. Data Integration Bus)

Business Architecture Sub-Domain

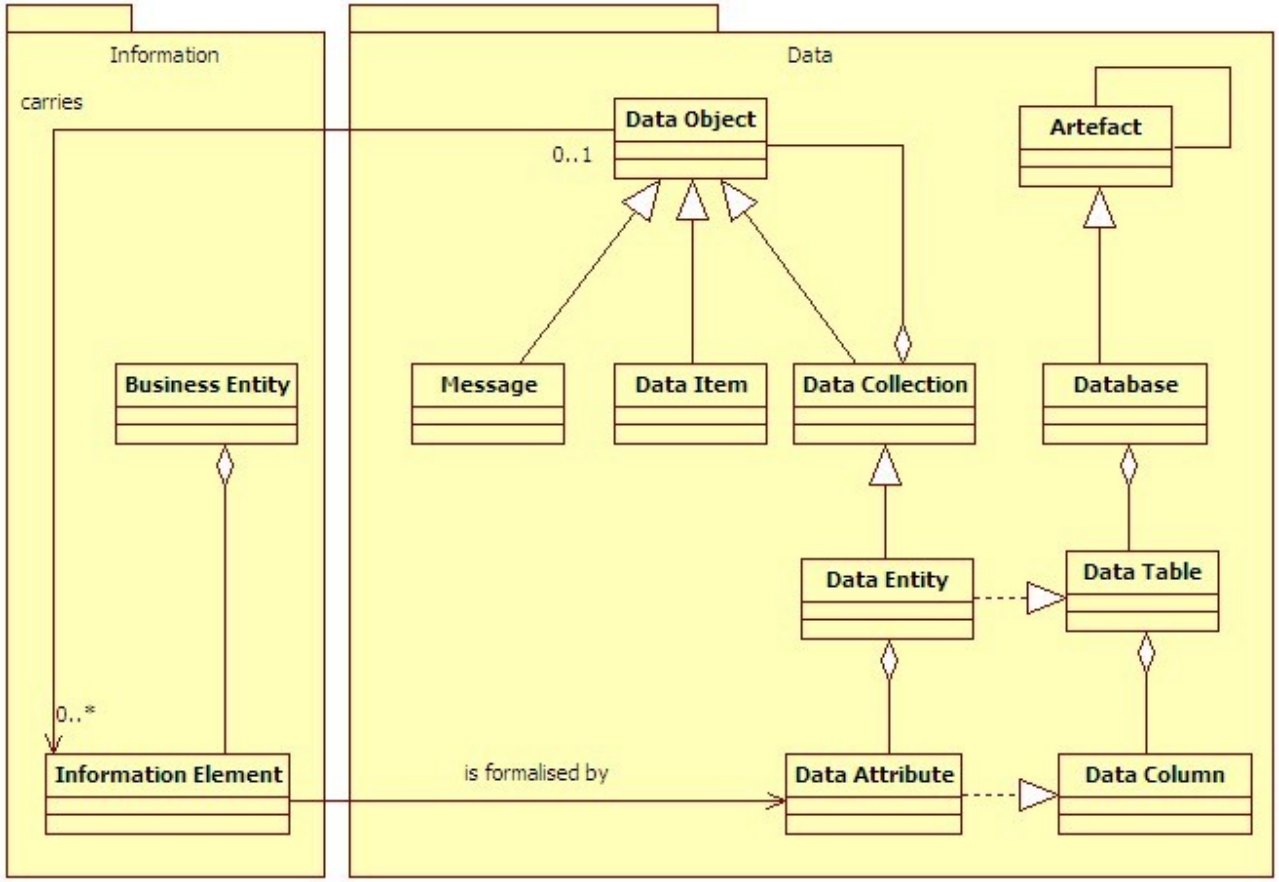


doriq

Business Architecture Context

- Key sub-domain for recording entity life-cycle events
- What happens to the data?
- How is it transformed?
- By whom?

Information Architecture Sub-Domain

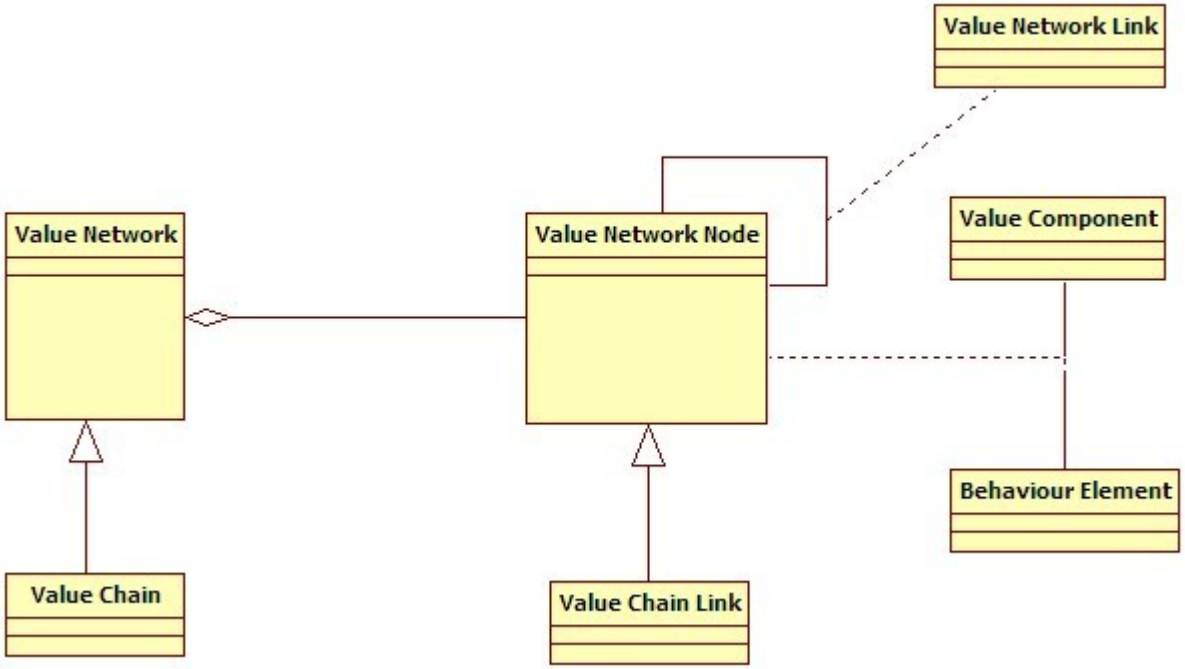


doriq

Information Architecture Context

- The core sub-domain for DG
- Provides formal metamodel for data
- Incorporates mapping between
  - conceptual
  - logical
  - Physical
- Relates structured data to compound data collections and unstructured data

Data Value Sub-Domain

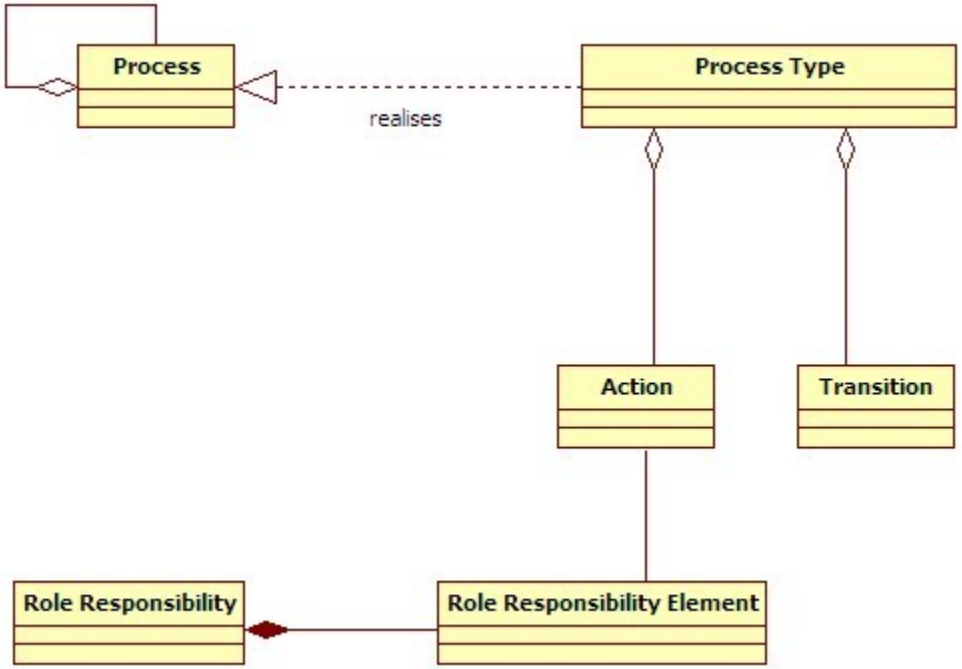


doriq

Data Value Context

- Provides formal structure for:
  - Value chains
  - Value networks
    - Related to data
- Value Components (that value provided by a Behavioural Element – an Activity) link back to Business Services
  - NB: not IS services!

Process Sub-Domain

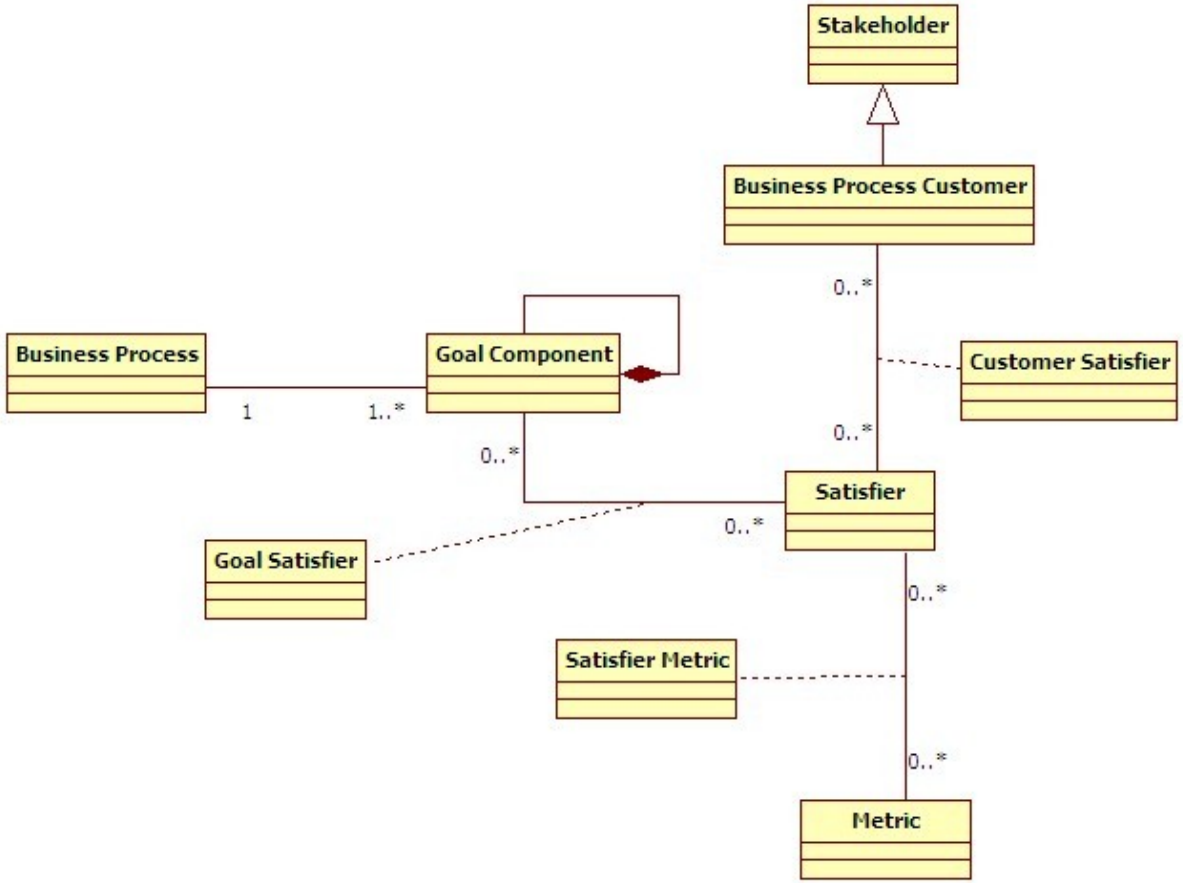


doriq

Process Context

- Provides further detail on the way in which roles may have responsibility:
  - For Data Stewardship
  - For ‘in-line’ data management during operational activity within the business

Measurement Sub-Domain



doriq

Measurement Context

- Show how business satisfiers address enterprise goals
- Also how these satisfiers are measured, and
- How they contribute to overall customer satisfaction (for the process customer)

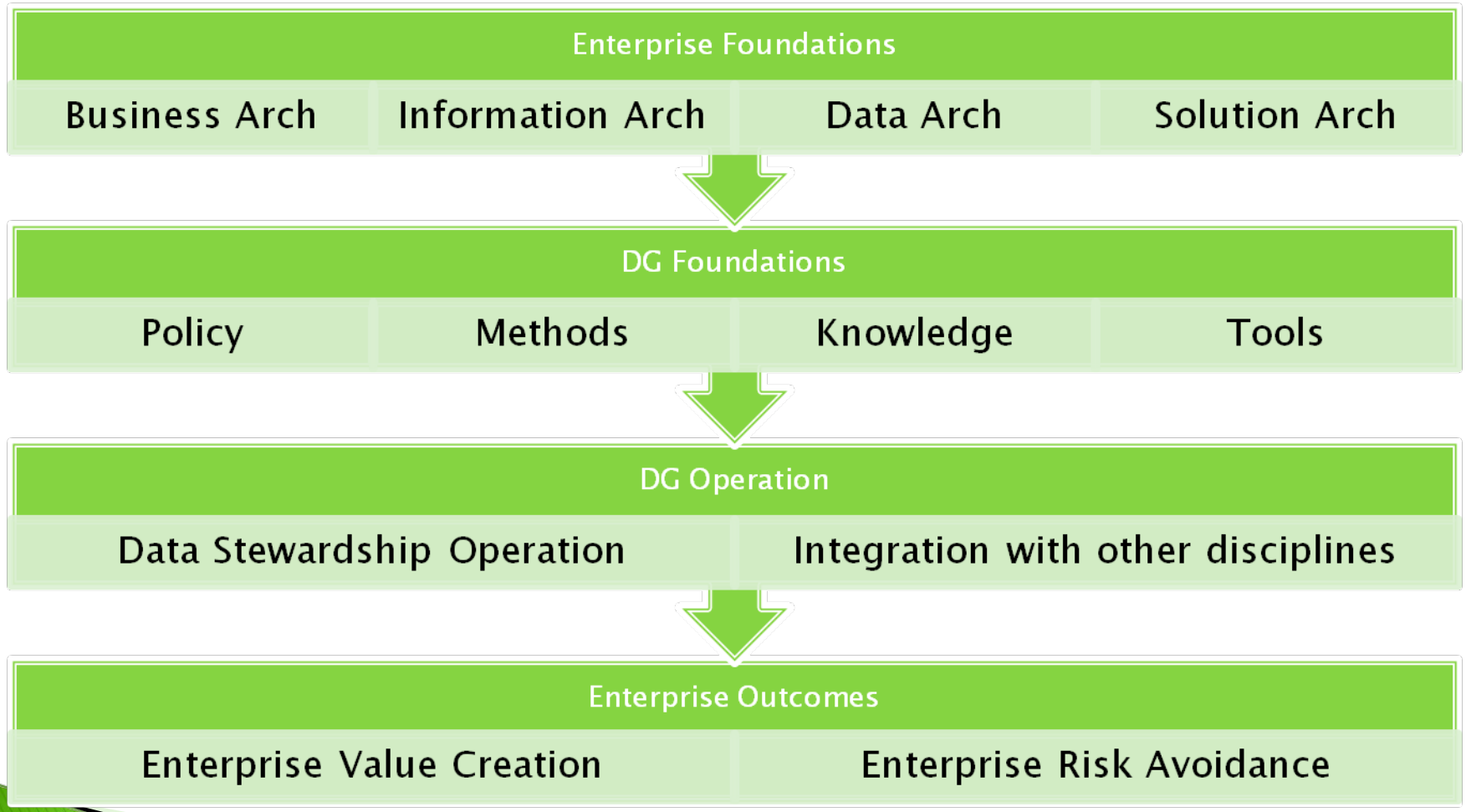
Sub-Domain Interrelationships

- Motivation domain provides context for setting Business Strategy
- Business Strategy provides context for development of Business Architecture
- Business Process domain describes how Business Strategy is achieved
- Business Capability describes how Business Processes are supported
- Value domain structures the elements that contribute to the strategy
- Measurement domain defines how these contributions are monitored over time
- Etc. Etc.

Part 4: Setting up Architected DG

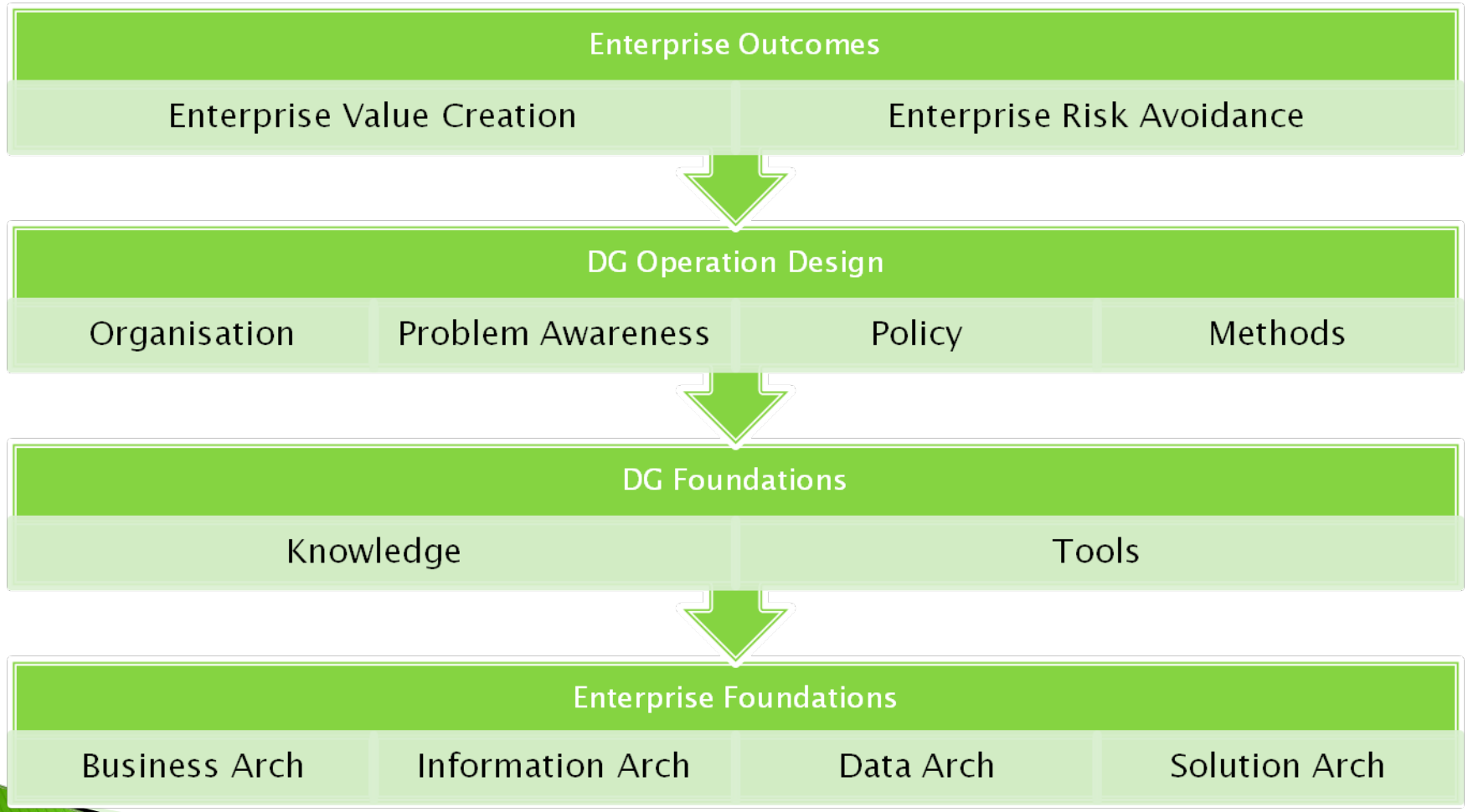
doriq

Architected DG Schematic (Operation)



doriq

Architected DG Schematic (Design)



doriq

Critical Success Factors

- Key Factors in successfully architecting a DG function:
  - Focus on Data Value at a strategic level (as determined by the business motivations/drivers)
  - Focus on Data Quality at the tactical level (based on fitness for purpose, as determined by the business context)
  - Structured common 'Knowledge Backbone'
  - Rigorous integration of DG at all levels: (Managerial, Architectural, Development, Operational)
  - Measurement of success based on:
    - Sustained/improved business value of data
    - Increased maturity in DG function

Types of Data Quality

- Intrinsic
  - Accuracy
  - Believability (*)
  - Reputation
  - Objectivity
- Contextual
  - Relevance (*)
  - Value Added (*)
  - Timeliness (*)
  - Completeness (*)
  - Depth of information (*)
- Representational
  - Interpretability (*)
  - Ease of Understanding (*)
  - Concise Representation
  - Consistent Representation
- Access-related
  - Accessibility
  - Security

(*) Truly CONTEXTUAL factors

The Case for Tight Coupling

- Data Governance must be tightly, accurately and reliably coupled with the business architecture
  - “Trying to understand the business domain by understanding individual data elements is like trying to understand a community by reading the phone book” – Ellen Friedman
- This is achieved through the context provided by the EBA for the EIA

Bridging Architecture and DG

- All truly Contextual (starred) categories of DQ problem must be defined and managed
  - These items require close coupling of Business and Information layers
- Data Governance must take into account all relevant interrogatives when defining and operating a data quality environment
  - Often highly focussed on WHAT, needing more emphasis on others (especially WHY)

Lack of an Enterprise Position

- Often, the DQM function is hampered in placing the right context on its activities by lack of an authoritative answer.
- This may occur because:
  - The EIA function does not exist
  - The EIA function is in fact just dealing with IS architecture (e.g. provision of Data Integration engines)
- NB: A sub-Enterprise level context will result in inefficiency, error and potentially failure of the DQM function

Coupling DG with Business Architecture

- Must ensure that business motivation is taken into account:
  - Drivers
  - Constraints
  - Rules
- Data management tasks should be seamlessly integrated with operational business processes
- Full life-cycle perspective must be defined for major entities in business activity context

Coupling DG with Information Architecture

- Must provide universal (enterprise) perspective on:
  - Data definition
  - Data life-cycle
  - Data value proposition
- Ensures that overall integrity is preserved within ‘segment’ activities

Coupling DG with Solution Architecture

- Data Availability across full life-cycle must be taken into account
- Facilities for
  - Stewardship
  - Life-cycle data transformations
  - Ad-hoc data availability
  - Consistent security and access management across enterprise

Coupling DG with Operational context

- Business operations
- IT Development
- IT Operations

Summary

- You can't solve the DG problem just by looking at data!
- The broader context includes understanding and control of:
  - People
    - Individuals and Organisations
    - Internal and External
  - Processes
  - Risk
  - Cost
  - Value, and the measurement of value

Conclusions

- Architected Data Governance provides:
  - An authoritative, enterprise-wide perspective
  - A sound basis for re-use of data, by ensuring that knowledge about data is:
    - Accurate
    - Trusted
    - Authoritative
    - Well-communicated
    - Embedded in existing practices
  - A solid basis for the measurement of DGs value itself, through linking with Maturity Modelling

Questions?

References

- Adler, Steven: "Six Steps to Data Governance Success" (http://www.cio.com/article/print/114750)
- Wang, R. Y. and Strong, D. "Beyond Accuracy: What Data Quality Means to Data Consumers."
  Journal of Management Information Systems, 12, 4 (Spring 1996), 5-34.
- Wang, R. Y.; Lee, Y. W.; Pipino, L; and Strong, D. M. "Manage Your Information as a Product."
  Sloan Management Review. 39,4 (Summer 1998), 95-105.
- IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance
- "Re-use and Reusability" – Presentation to Enterprise Architecture Europe, 2007 (http://www.doriq.co.uk/downloads2/cat_view/53-presentations)
- "Quality Data Through Enterprise Information Architecture", Semyon Axelrod (http://msdn.microsoft.com/en-us/arcjournal/bb266338.aspx)

Contact Information

- Guy Tozer
- Doriq Associates

- guy.tozer@doriq.co.uk
- www.doriq.co.uk

- ++44 (0)7847 303 176

Thank you for your attention!

doriq