The MIT 2008 Information Quality Industry Symposium

Patterns in Data Quality
A Method for Organizing
Enterprise Data Quality (Web) Services
in Service Oriented Architectures

Michael Overturf – VP of Strategy

Navin Sharma – Dir. of Product Management

The MIT 2008 Information Quality Industry Symposium

Patterns in Data Quality

- A methodical approach to structuring rules for data user satisfaction
- Patterns provide simplification
- Patterns focus measurement
- System architects are the primary beneficiaries of simplification

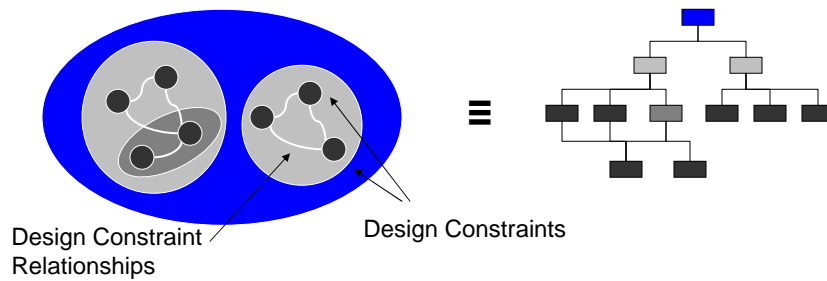The MIT 2008 Information Quality Industry Symposium

Background

Structured constraint patterns developed by Christopher Alexander (Architect and Mathematician) in 1964 in '*Notes on the Synthesis of Form*'

M – Set of Design Constraints
L – Set of Design Constrain Relationships

G(M,L) – Linear Graph of Design Constraints

≡

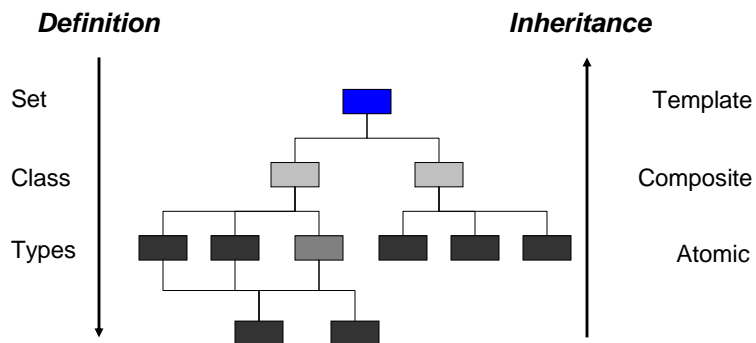Design Constraint Relationships

Design Constraints

The MIT 2008 Information Quality Industry Symposium

Constraint Tree Dynamics

Definition-Inheritance dynamic describes how we get from DQ Constraint Set to DQ Service

Definition   *Inheritance*

Set   Template
Class   Composite
Types   Atomic

The MIT 2008 Information Quality Industry Symposium

Constraint Patterns in Data Quality

*Data quality constraints
define
the form of data quality rules*

Metadata	*Platform*	*Access*	*Modality*	*Input*	*Output*
Consistency	Operating System	Web Services/SOAP	Transactional	Type	Type
Completeness	SaaS	C/S, API	Batch	Binding	Binding
Accuracy		Security	Microbatch	Source	Sink
Uniqueness/ Singularity			Multi-modal (Any)		

The MIT 2008 Information Quality Industry Symposium

Pattern yields Form
A simple example

Constraint Pattern
'Flurdleneep'

Accuracy — Input — Output — Modality

Normalize Name — Corporate Name — Name_Field — Name_Field — Multi

Name Parsing Function — Corporate Name Structure — Name_Field Input Map — Name_Field Output Map

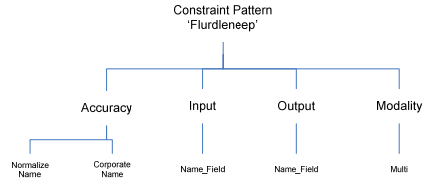Parser — Input — Output — Modality

Transactional or Batch Subflow

The MIT 2008 Information Quality Industry Symposium

Pattern yields Form
A simple example

Constraint Pattern
'Flurdleneep'

Accuracy   Input   Output   Modality

Normalize Name   Corporate Name   Name_Field   Name_Field   Multi

Subflow

Input   Name Parser   Output

Free Input   *Deployed Parser*   *Free Output*
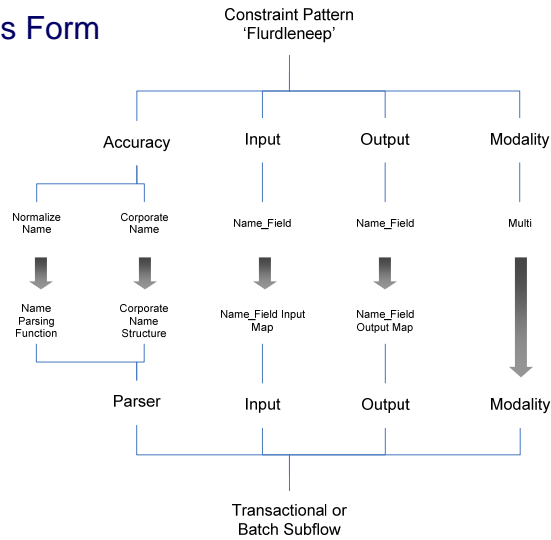
The MIT 2008 Information Quality Industry Symposium
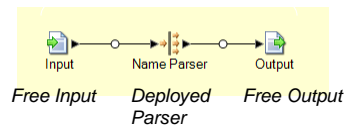
Quality as a Service (QaaS)

Patterns naturally translate into services (in a SOA context):

– *Atomic* services that are fine grained to provide a monotonic function;

– *Composite* services that encompass two or more atomic services;

– *Templates* make up one or many composite and atomic services
    – by specific data quality metrics (consistency, uniqueness, etc.)
    – industry specific business processes

Template

Composite

Atomic

The MIT 2008 Information Quality Industry Symposium

Example:
Insurance
Underwriting

Business Process

Template

Composite Service

Atomic Services
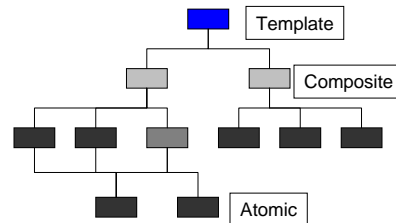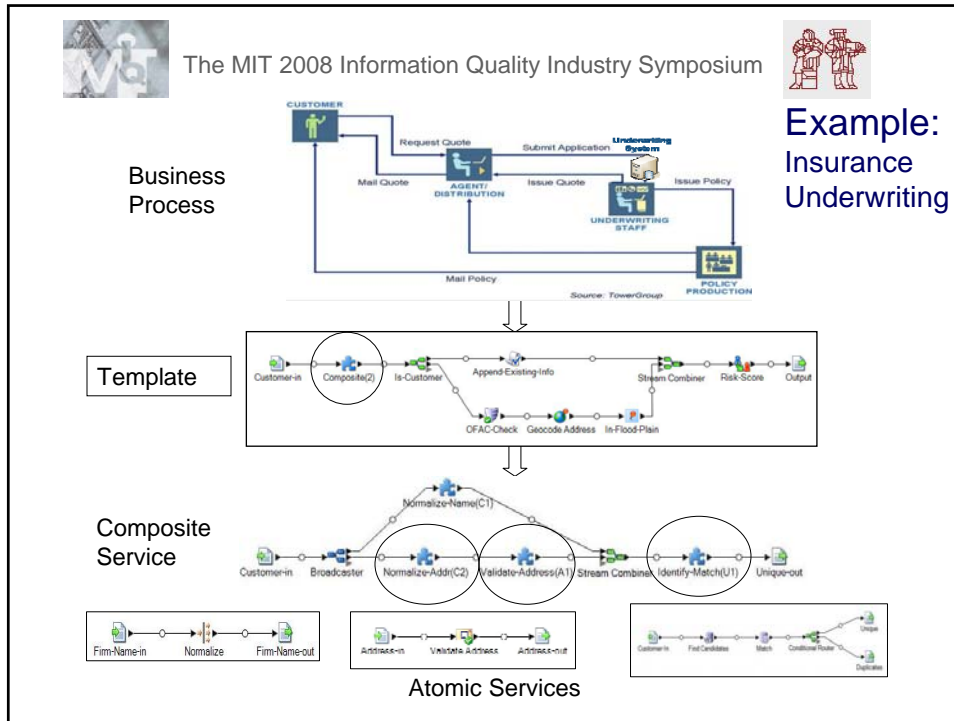
The MIT 2008 Information Quality Industry Symposium

Web Services implement Pattern Constraints

- Widely adopted and accepted within enterprises and vendor communities
- Software community agreement on standards for distribution
  - WS-I and W(3)C standards,
  - WSDL (Web Services Description Language)
- Principles encourage and enable –
  - Ease of integration (loose coupling)
  - Re-use
  - Ease of management
  - Agility
  - Active use
- Satisfies modality constraints (batch, transactional)

The MIT 2008 Information Quality Industry Symposium

Laying the foundation

- **Create** – a Composition Interface that allows data stewards and enterprise architects to use and define templates, composite and atomic data quality services
- **Publish** – for distribution
- **Integrate** – via standards based interfaces (WSDL, etc.)
- **Manage** – for governance

Integrate

Orchestrate

Develop

Patterns

Secure

Analyze

Access

Manage
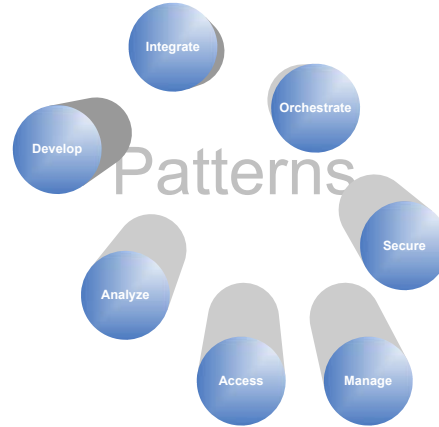
Consistent Measurement of Data Quality Results

The MIT 2008 Information Quality Industry Symposium

Summary

- A methodical approach to structuring rules for data user satisfaction
- Patterns provide simplification
- Patterns provide a structure for measurement of data quality
- System architects manage data quality using standard Web Service Management Lifecycle

The MIT 2008 Information Quality Industry Symposium

Contact Information

- Michael Overturf
  - Email: Michael_Overturf@g1.com
  - Phone:413-695-5500

- Navin Sharma
  - Email: Navin_Sharma@g1.com
  - Phone: 240-447-6801