



# Customer Data Quality/Data Integration

## What's New

Jay Bourland  
Pitney Bowes Business Insight







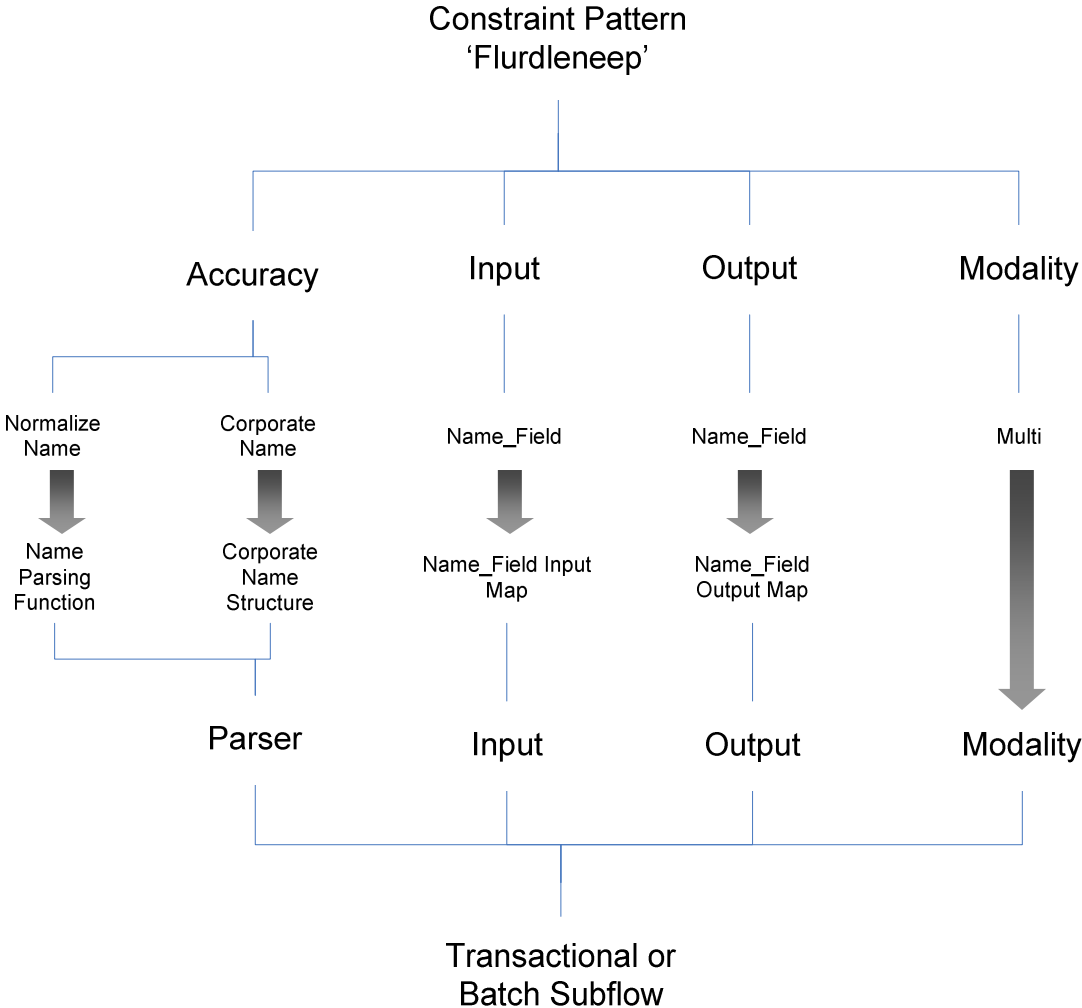
# Constraint Patterns in Data Quality

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

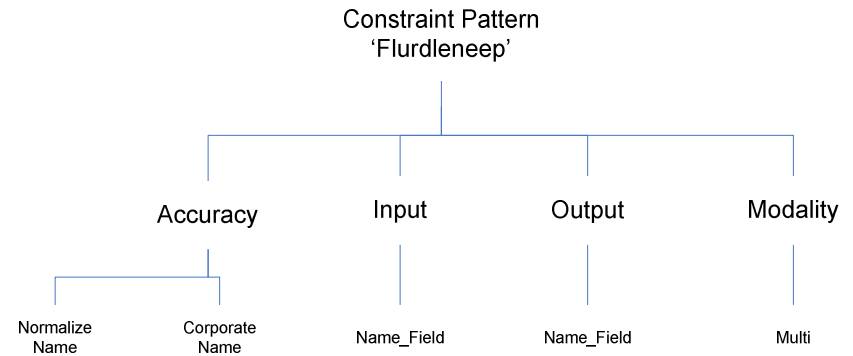
<u>Metadata</u>	<u>Platform</u>	<u>Access</u>	<u>Modality</u>	<u>Input</u>	<u>Output</u>
Consistency	Operating System	Web Services/SOAP	Transactional	Type	Type
Completeness	SaaS	C/S, API	Batch	Binding	Binding
Accuracy		Security	Microbatch	Source	Sink
Uniqueness/ Singularity			Multi-modal (Any)		



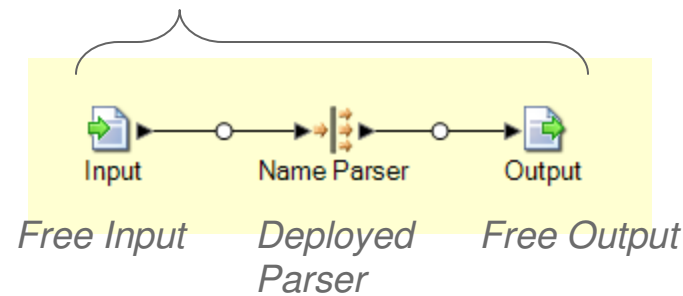
# Pattern yields Form – A simple example



# Pattern yields Form – A simple example



## Subflow



# Quality as a Service (QaaS)

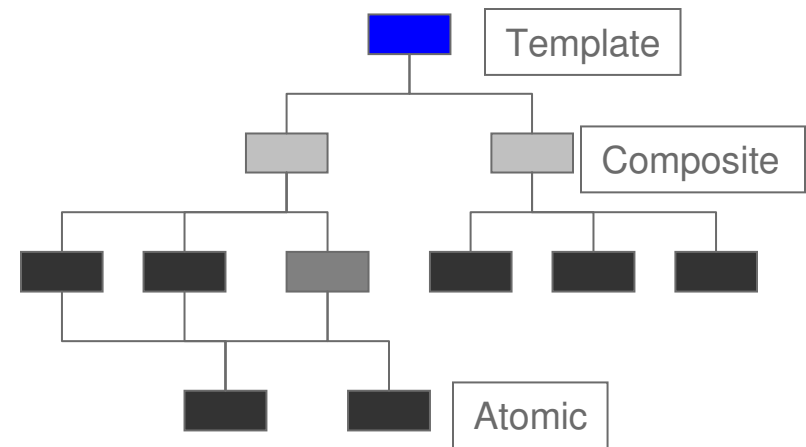
Patterns naturally translate into services (in a SOA context):

*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





## 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)





# 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







# Data Flow 3 Year Roadmap

- Development Enhancements
  - Dev to Test To Production Migration
  - Plan Versioning
  - Improved User Interface
- Expanded Connectivity
  - Improved Mainframe and ERP/CRM Connectivity
- Administrative Improvements
  - Thin Administrative Interface
  - Multiple Administrators with Role Rights
  - Unified Plan Execution Tracking
  - Improved Plan Tracking
  - Improved MetaLens
  - Clustering Support







# CDQP Release November 28<sup>th</sup>, 2008

- Expand SaaS offering
  - Batch support
  - Mixed data flows including local and hosted (SaaS) stages
- New features for increasing international support
  - Open parser to support all Country, Region and Languages
  - Open, customize and prioritize Universal Addressing Module data sources based on country
  - Plug in Matching logic for support of international keyboards, phonetics, etc...
- Deeper enterprise customer offering
  - Support for multiple CDQP Servers
  - Hot swap of databases
  - Dataflow versioning, comments and promotion from development to test to deployment
- Enhanced Data Steward Capabilities





# Increased International Support

The screenshot displays the 'Domain Editor' application window. The 'Cultures' tab is active, showing a list of cultures including Arabic - Iraq (ar-IQ), Arabic - Jordan (ar-JD), English - United States (en-US), English - United Kingdom (en-UK), English - Canada (en-CA), German - Germany (de-DE), Spanish - Mexico (es-MX), and Spanish - Spain (es-ES). The 'Personal Names' domain is selected. The main text area contains the following code:

```
%Tokenize([:space:]);  
%InputField("Name");  
%OutputFields("FirstName","MiddleName","LastName");  
  
<root> = <FirstName><MiddleName><LastName> |  
        <FirstName><LastName>;
```

Below the code is a checkbox for 'Enable word wrap' and a 'Clear' button. A 'Test data' field contains 'John Adam Smith' and an 'Update' button. At the bottom, there are 'Add...' and 'Remove' buttons. On the right side, there is a 'Commands' panel with a table of commands and their descriptions.

Command	Description
%Tokenize(character set[...])	Defines what
%InputField("name")	Defines the in
%OutputFields("name"[, "h...	Defines the c
%Culture("id")	Defines what
@RegEx("expression")	Matches a to
@Table("table-name")	Matches a to
@Table("table-name", "exp...	Matches a to
@Token	Matches any
(expression)	Wraps the se
	Indicates a c
?	Expression m









# Questions? Suggestions?

