Data Quality Services.

Making Data Fit For Business

Copper Blue Consulting

Who am I?

- Allan Mitchell
- Joint author on 2005/2008 SSIS Book by Wrox
- Websites
  - www.CopperBlueConsulting.com
- StreamInsight and SQL Azure Advisory Boards
- Specialise in Data and Process Integration
- Microsoft SQL Server MVP
- Twitter: allanSQLIS
- E: allan.mitchell@CopperBlueConsulting.com

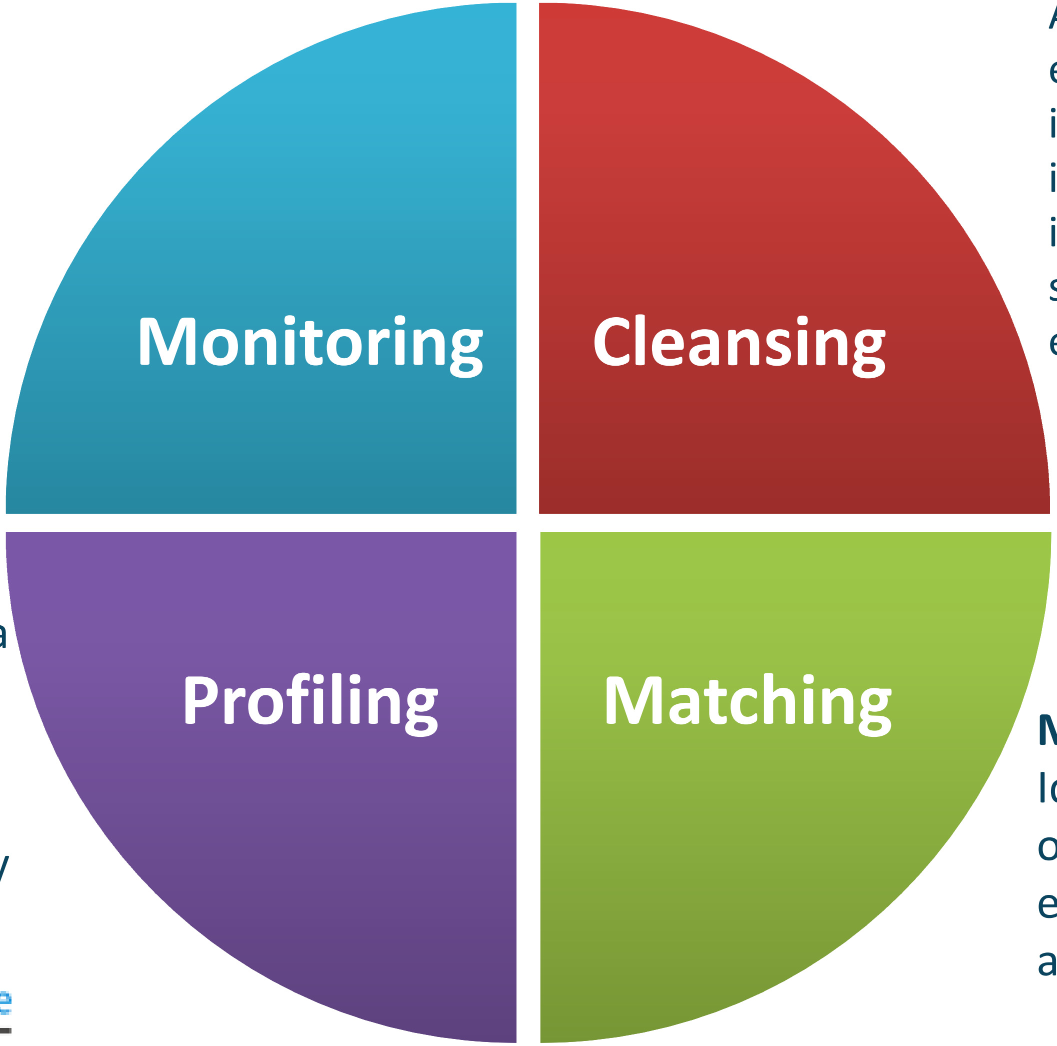
Copper Blue
Consulting

Common Data Quality Issues

Data Quality	Issue	Sample Data Problem
Standard	Are data elements consistently defined and understood ?	Gender code = M, F, U in one system and Gender code = 0, 1, 2 in another system
Complete	Is all necessary data present ?	20% of customers' last name is blank, 50% of zip-codes are 99999
Accurate	Does the data accurately represent reality or a verifiable source?	A Supplier is listed as 'Active' but went out of business six years ago
Valid	Do data values fall within acceptable ranges?	Salary values should be between 60,000-120,000
Unique	Data appears several times	Both John Ryan and Jack Ryan appear in the system – are they the same person?

Requirements for Data Quality Solutions

Monitoring
Tracking and monitoring the state of Quality activities and Quality of Data



Cleansing
Amend, remove or enrich data that is incorrect or incomplete. This includes correction, standardization and enrichment.

Profiling
Analysis of the data source to provide insight into the quality of the data and help to identify data quality issues.

Matching
Identifying, linking or merging related entries within or across sets of data.

Copper Blue Consulting

What is DQS ?

Data Quality Services (DQS) is a **Knowledge-Driven data quality solution,** enabling IT Pros and data stewards to easily improve the quality of their data

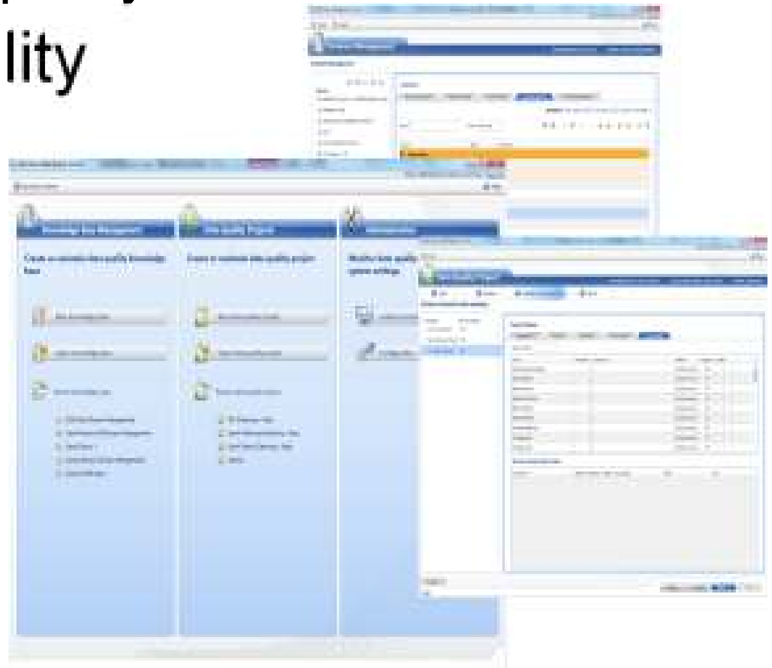
Copper Blue Consulting

Microsoft's DQS Solution Concepts

Knowledge-Driven	• Based on a **Data Quality Knowledge Base** (DQKB)
Semantics	• **Data Domains** capture the **semantics** of your data
Knowledge Discovery	• **Acquires additional knowledge** the more you use it
Open and Extendible	• Support use of **user-generated knowledge** and IP by 3rd party **reference data providers**
Easy to use	• Compelling user experience designed for **increased productivity**

Make Data Quality Approachable To Everyone

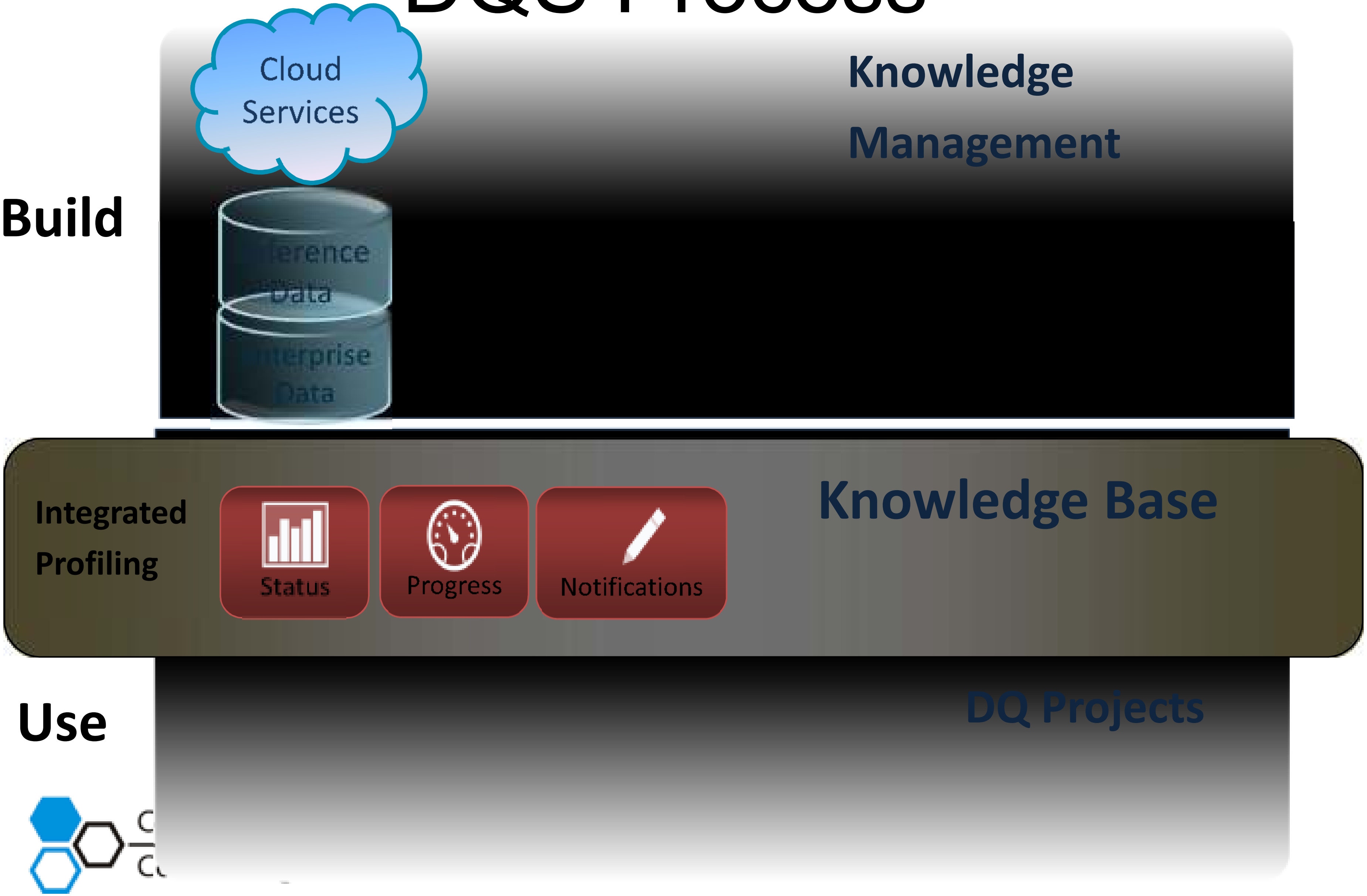
- Improve your data quality with DQS
  - Cleanse the data and keep it clean
  - Build confidence in your enterprise data
  - Share the responsibility for data quality
- Remove Barriers for Data Quality
  - Designed for ease of use
  - Empowering the business users
  - See data quality results in minutes rather than months



Copper Blue
Consulting

Demo: DQS UI

DQS Process



Build

Cloud Services

ference Data

terprise Data

Knowledge Management

Integrated Profiling

Status

Progress

Notifications

Knowledge Base

Use

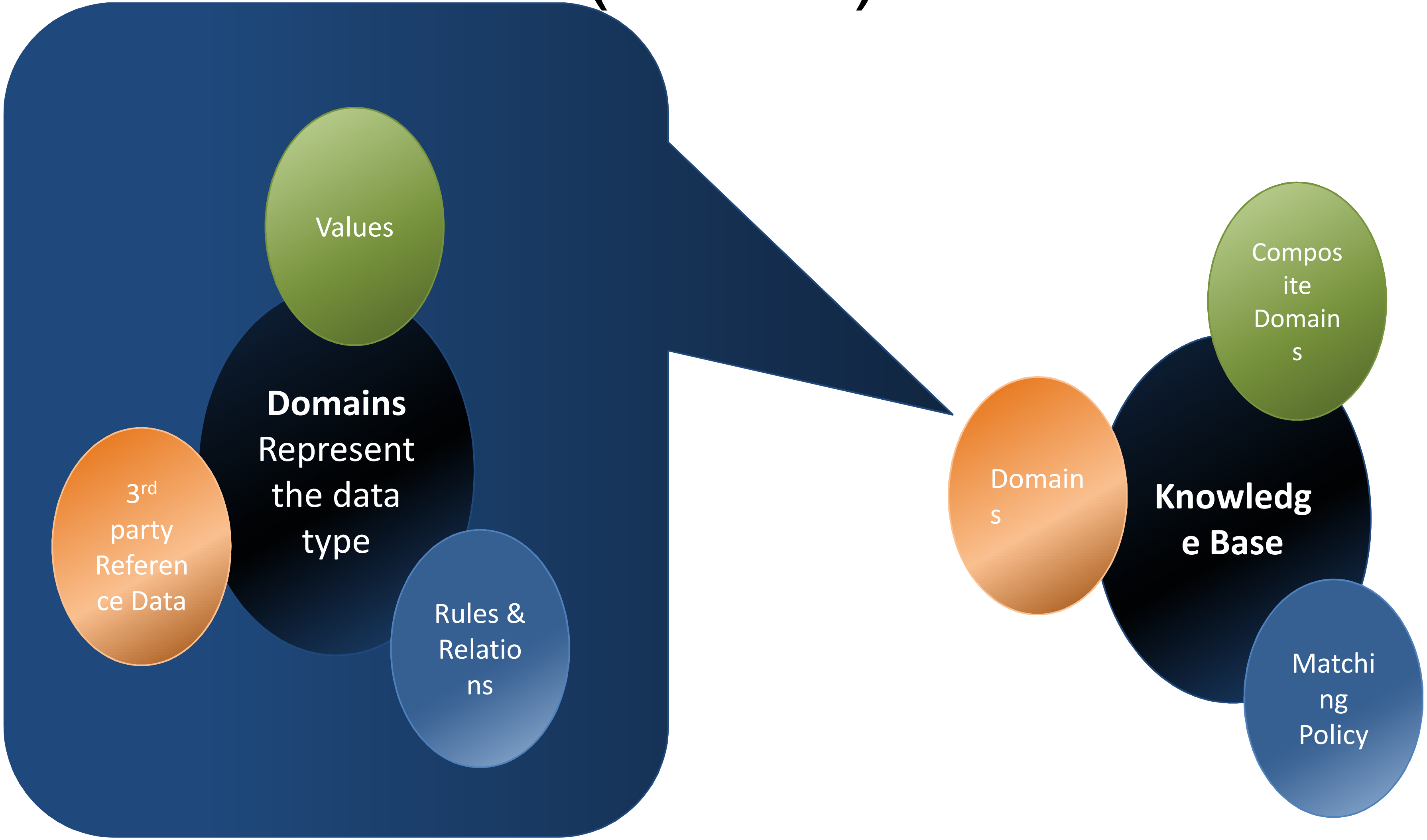
DQ Projects

DQS High Level Scenarios

Knowledge Management & Reference Data	• Creating and managing the Data Quality Knowledge Bases • Discover knowledge from your org’s data samples • Exploration and integration with 3rd party reference data
Cleansing & Matching	• Correction, de-duplication and standardization of the data
Administration	• Tools to monitor and control data quality processes

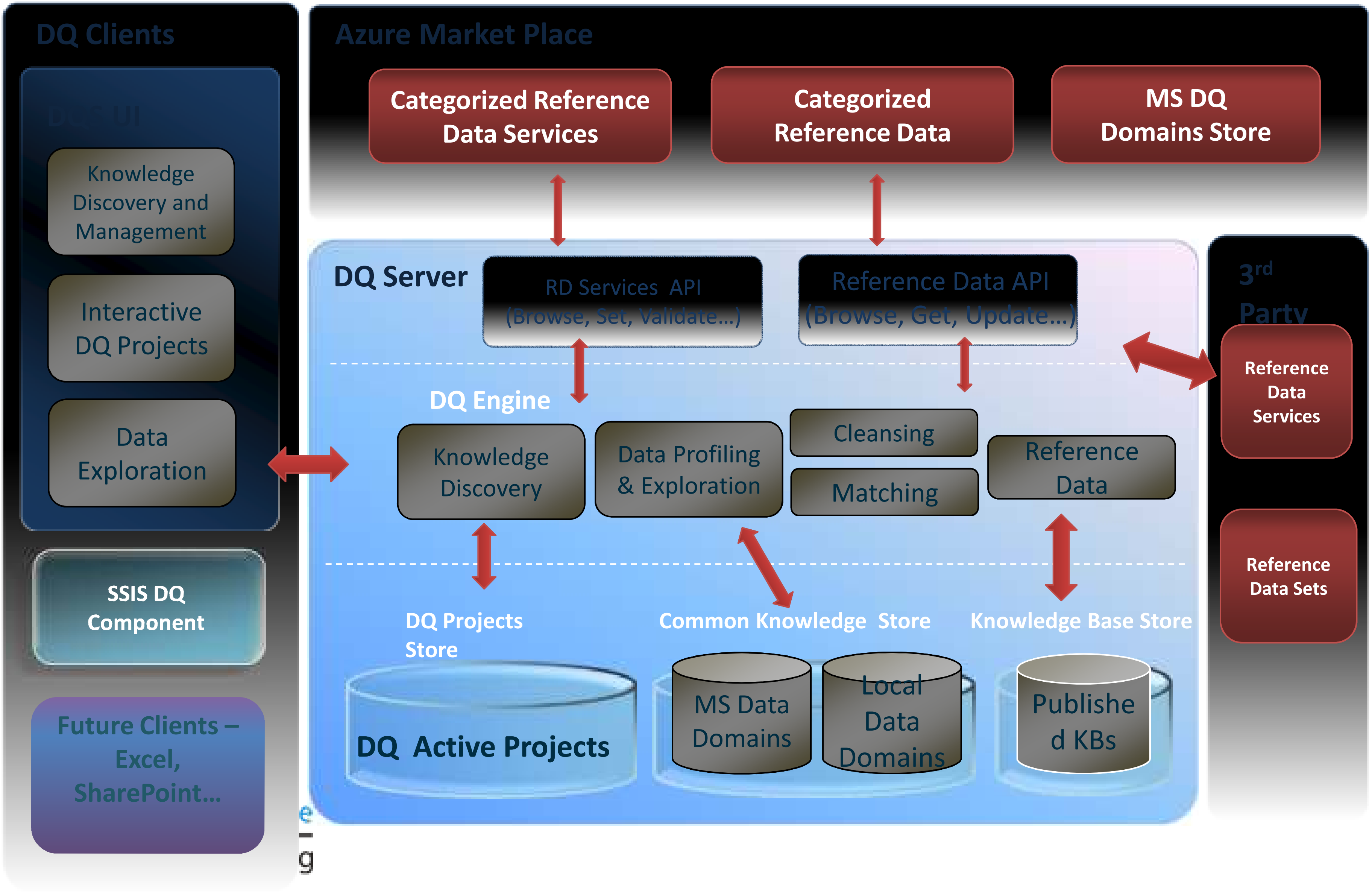
Copper Blue Consulting

Data Quality Knowledge Base (DQKB)



Copper Blue Consulting

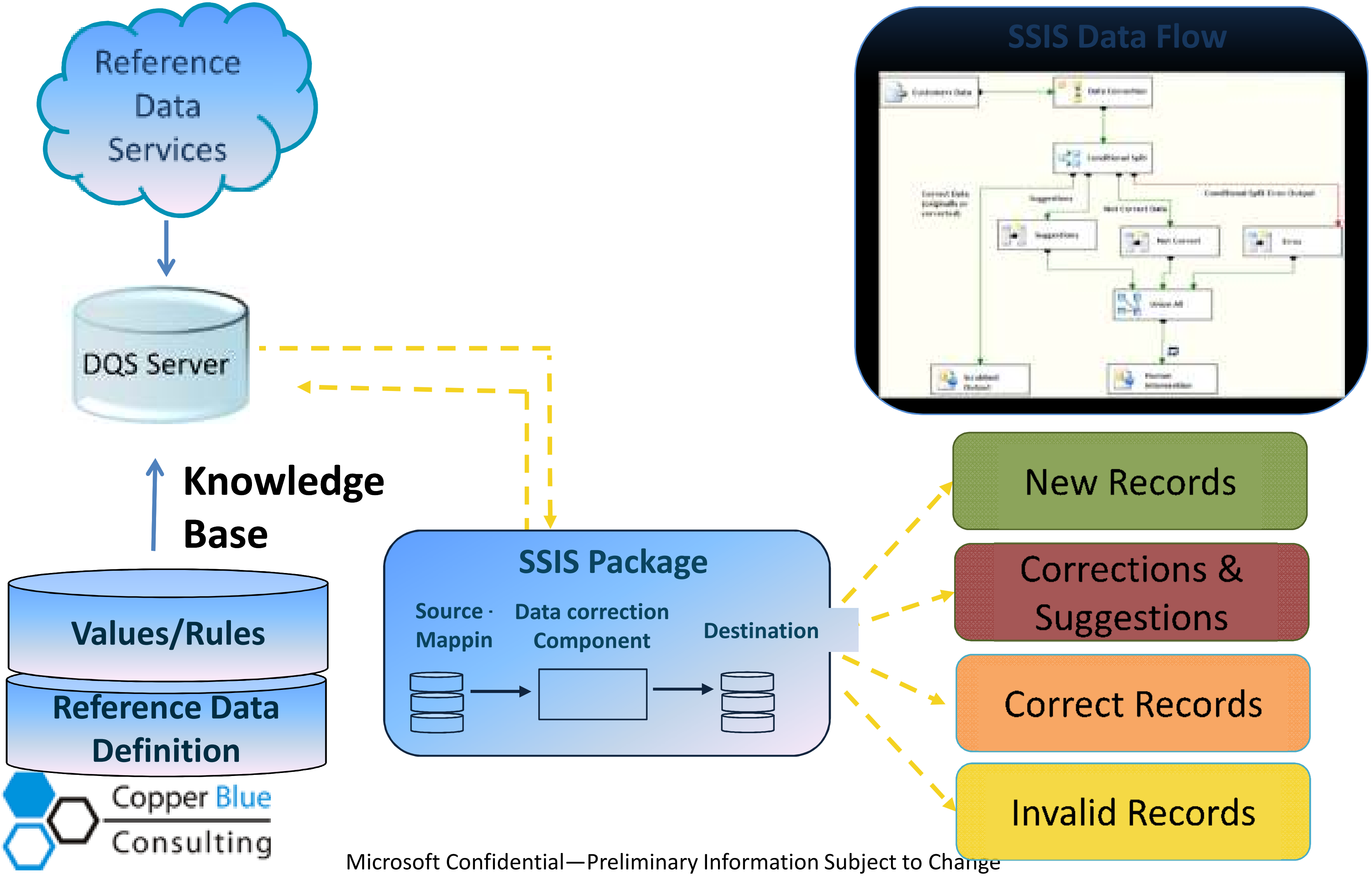
DQS Architecture Overview



DQS Data Sources

DataMarket	Easily cleanse and enrich data with Reference Data Services from DataMarket
3rd Party Reference Data Providers	Open integration with external 3rd party reference data providers
DQS Data Store	Website that contains DQS knowledge available for downloading
Organization Data	Create domains from your own data sources
Out of the Box Knowledge	A set of data domains that come out of the box with DQS

Batch Cleansing - Using SSIS



Demo: DQS Cleansing

Copper Blue Consulting

Matching

- Why Match?
  - Identify duplicates within the data source
  - Create consolidated view of data
- DQS Matching
  - Build a matching policy
  - Matching training
  - Create a matching project
  - Choose survivors

Demo: DQS Matching

Copper Blue Consulting

DQS – Value Proposition Summary

Knowledge-driven

- Rich Knowledge Base
- Continuous improvement and knowledge acquisition
- Build once, reuse for multiple DQ improvements

Easy To Use

- Focus on productivity and user experience
- Designed for business users
- Out-of-the-box knowledge

Open & Extendible

- Focus on cloud-based Reference Data
- User-generated knowledge
- Integration with SSIS

Bringing it all together

demo