

## TECHNOLOGY AUDIT

# Customer Data Quality Platform









## Pitney Bowes Business Insight




### BUTLER GROUP VIEW

#### ABSTRACT

*Pitney Bowes Customer Data Quality Platform (CDQP) is a domain-specific customer data quality management system that enables tasks such as integration, cleansing, matching, profiling, monitoring, and enriching the data with location and demographic information. For most B2C companies, accurate customer data and a single consolidated view of the customer is crucial for efficient outreach programs, efficient decision making, and also regulatory compliance for certain sectors such as banking. CDQP's core operations include parsing complex data types such as names and addresses into structured metadata, standardising these elements, normalising the data, and finally matching records to group them under common names and other entities. The solution is aimed at the data steward/data management specialist, who can configure matching rules and formulate dataflows, encompassing their business rules with CDQP core services that can be executed in batch or real-time mode. All of CDQP's data cleansing and quality management capabilities are available as services to be used against any set of customer information, and assembled into dataflows. Overall, Butler Group believes that CDQP is a strong product that would be useful in data governance initiatives at B2C companies.*

#### KEY FINDINGS

- |   |  |
|---|--|
|  Offers a wide range of integration options, including Web services.         |  Workflows can be designed for data quality operations which can be reused as services. |
|  Address variations for 250 countries and name recognition for 438 cultures. |  Comprehensive matching engine helps discover relationships between records.            |
|  No role-based security implemented within the platform.                     |  Specialised modules for Siebel, mySAP, and Microsoft Dynamics CRM.                     |
|  Partners with Master Data Management vendors.                               |  Available both as hosted and Software-as-a-Service.                                    |

Key:  Product Strength  Product Weakness  Point of Information

#### LOOK AHEAD

The pipeline includes: tighter integration with profiling and monitoring tools and expanded international geocoding.

## FUNCTIONALITY

Most large Business-to-Customer (B2C) companies have siloed data management practices, which leads to the lack of a single or even a logically grouped view of the customer and a high frequency of erroneous data. Naturally, this leads to inefficient direct outreach programs (and inefficient analytics and decision making in general), reputational risks, and for banks with stringent Know Your Customer (KYC) norms, severe regulatory risks. These companies would benefit from a central logical layer of data, built and maintained collaboratively by a multitude of stakeholders. The obstacles are of course less technical than sociological. However, such data governance and Customer Data Integration (CDI) initiatives would benefit from a data quality management solution that is 1) Domain specific, given the need for quick results for organisation-wide acceptance and the specialised nature of customer information; and 2) Flexible enough to be quickly integrated with a set of information and analytics, Line of Business (LOB) applications, and flexible enough to quickly cleanse, standardise, and match customer data without long drawn-out IT projects.

### *Product Analysis*

Pitney Bowes Business Insight's (PBBI) Customer Data Quality Platform (CDQP) aims to address the aforementioned challenge. The solution parses customer data into components such as name and address, cleanses, standardises, and normalises the data (based on established standards in the customer information domain), and identifies duplicate records and relationships across records based on a deterministic and probabilistic approach that incorporates a combination of business rules and fuzzy algorithms. Crucial initial steps preceding the aforementioned tasks would be integrating data from multiple sources and profiling customer records, a process that yields a summary of the quality, fields, and joins in the data. These processes could be followed by data enrichment, and customer address information could be geographically indexed to add demographic and location-specific intelligence to the customer record. The solution is available through a series of separately licensable modules – including addressing module, names module, normalisation module, matching module, geocoding (or location indexing) module, profiling module, a tax module, and a monitoring module, among others. CDQP also comes bundled in business application-specific modules, for Siebel Enterprise Apps, Siebel Industry Apps, mySAP Customer Relationship Management (CRM), mySAP Enterprise Resource Planning (ERP), and Microsoft Dynamics CRM. The solution is available hosted as well. Butler Group believes that the following aspects of the CDQP solution merit special mention:

**The integration and dataflow capabilities:** Butler Group is particularly impressed with the solution's focus on easy integration with LOB applications and the flexibility it affords for quick deployment of data quality processes. The entire set of core tasks mentioned at the beginning of this section is available as individual services and accessible through Application Programming Interfaces (APIs) and Web services to the existing Extract, Transform, and Load (ETL) infrastructure and enterprise applications. These services (which can operate in real time or in the batch mode) can be assembled into composite services, which can be reused. This reuse capability lends itself to a strong Return on Investment (ROI) for customers seeking rapid deployment when new requirements arise in the future.

**The matching engine:** enables the linking of records through deterministic and probabilistic methods and a typical use case would be grouping members of a household based on a common address and linking multiple records of an individual through a date of birth. The solution ships with matching rule sets (and PBBI's pipeline includes plans to increase the range).

**Comprehensive coverage of regions and culture:** Pitney Bowes is a veteran in the direct mailer and location intelligence market, and has provided address standardisation solutions internationally for two decades. CDQP supports Address validation for 250 countries and name recognition for 438 cultures.

Overall, Butler Group is impressed with the solution and believes that CDQP would be a crucial tool for companies initiating data governance or CDI programs. It is important to mention here that Pitney Bowes is completely focused on customer or party domain, and the solution would not be suitable for other data types, in spite of the technical capabilities of its core components, such as the matching engine.

## *Product Operation*

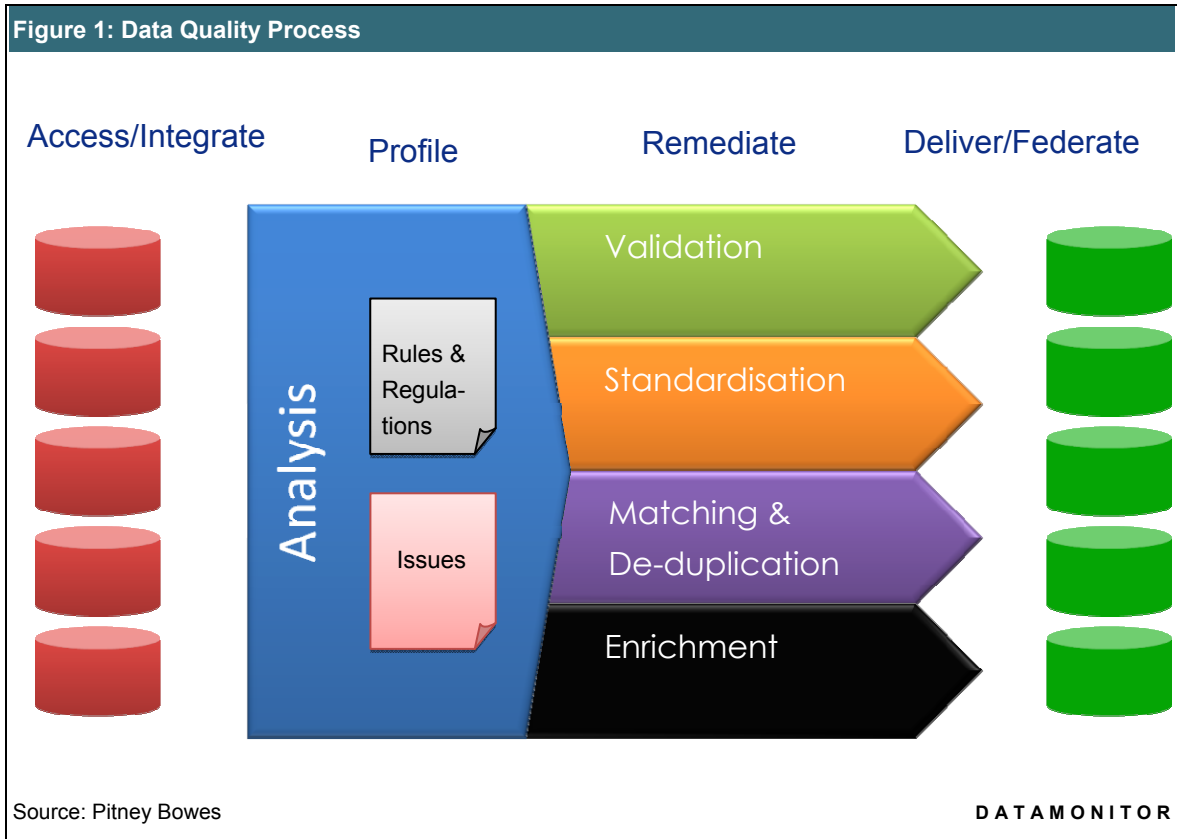
Pitney Bowes' CDQP comprises the data quality services integration, profiling, cleansing, matching, enrichment, and monitoring which can be used as services and assembled into composite services, and used both in the real-time and batch mode. The following list details some of CDQP's core set of services:

**Cleansing:** enables parsing, standardisation, and validation of data against reference data such as name, and address. Cleansing is in effect a supervised learning process and is GUI-driven, with the data specialist modifying terms and patterns to improve the parsing process. Special mention needs to be made of the variety of reference data types CDQP provides. The solution ships with standard libraries of nicknames, titles, and company names (with the many known variations of each entry). Pitney Bowes states that names from 193 cultures are supported. The range of cultures supported would be a useful and potentially even an essential capability, given the complex demographic shifts in developed markets.

**Matching/Consolidation:** enables identification of unique and duplicate, records; enables consolidation of records of individual customers into households, company accounts, and multiple records of a single individual by product (among other possibilities). The identification and linking is performed through GUI-driven rule sets which use key data items such as product, code, date of birth, and telephone number. Rules can be reused; matches resulting from different rule sets can be compared; and data across grouped data can be aggregated.

**Enrichment:** The consolidated and rationalised data set can be enriched using user-provided and/or third-party data. Common sets of data include gender, ethnicity, demographics, and geographic features (such as distance from coastline). CDQP can use Pitney Bowes' solution MapInfo to append geocoding information for US, Canada, Australia, the UK, and 20 other countries.

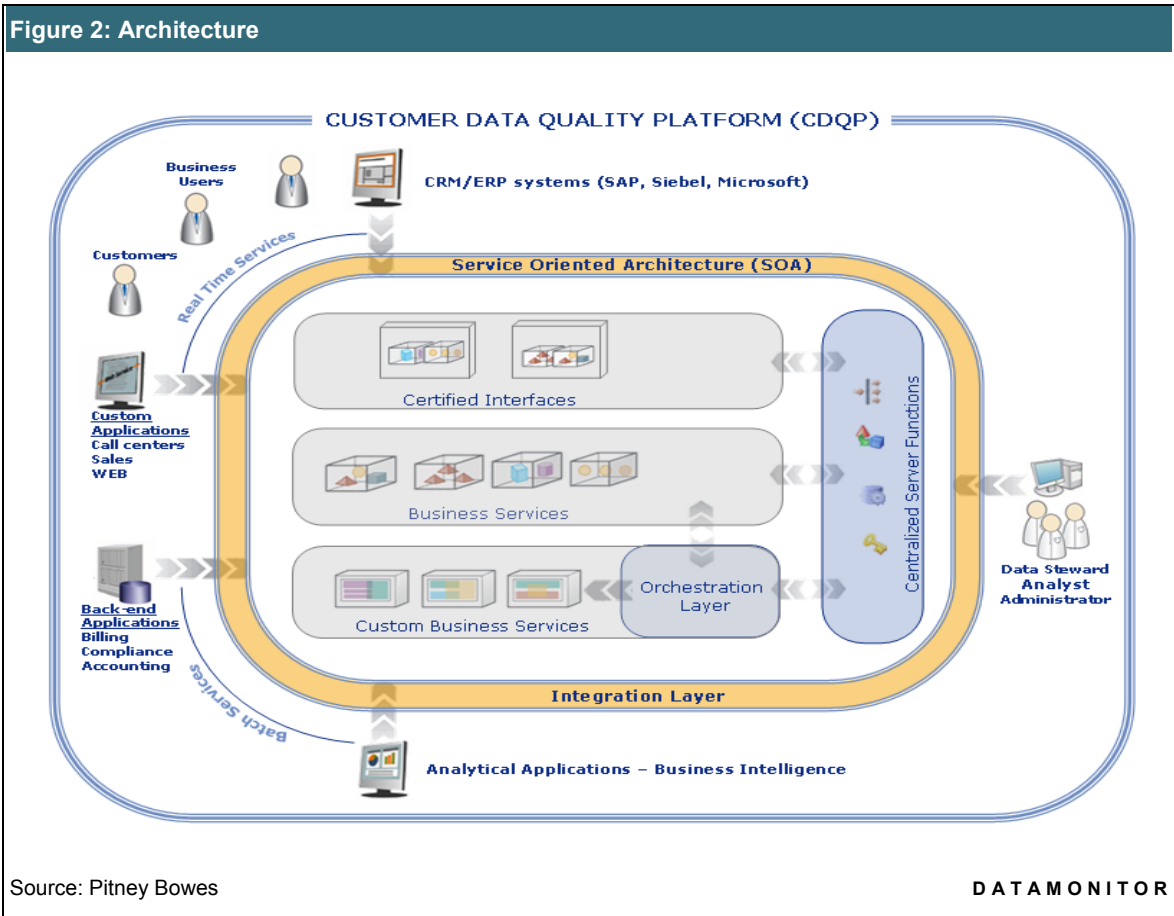
**Monitoring:** enables continuous monitoring of data quality, using a 'what-if' feature, alerts based on customisable rules, reports, and dash-boards. This is a key point in that while the data stewards might define and publish the business rules, the solution delivers value across the organisation by monitoring customer data and alerting users up and down the organisation of issues and trends.



The aforementioned services can be used by data stewards from within LOB applications or using CDQP clients through a set of integration capabilities and an orchestration layer that allows data quality services to be assembled, executed in batch mode and real time, and reused. Using the CDQP client Software Development Kit (SDK), integrations can be built using C, C++, COM, .NET, Java, Web services, and Extensible Markup Language (XML). Using Web services-based integration, and composite data quality services can be published (a Web Services Description Language (WSDL)) with a single click. In addition, certified connectors are provided for a host of common business applications, including SAP (R3, ERP Central Component, Customer Relationship Management (CRM), and vertical solutions), Siebel (including industry applications), and Microsoft Dynamics CRM.

The CDQP solution comprises the Enterprise server (the Java-based process management component), and a number of .NET-based thick-client tools. The Enterprise Designer is used by data stewards to create composite dataflows and has a drag-and-drop visual environment. Another key client is the 'Management Console', the administration client for services configuration, databases, Java Database Connectivity (JDBC) driver and remote file configurations, and reporting and logging operations.

Figure 2: Architecture



Pitney Bowes sets the performance benchmark for all functionality at a minimum of 1M/hr in batch process on a two-processor box, and sub-second in a real-time, multi-user, concurrent processing environment. The solution is scalable both vertically and horizontally. The server can scale through a load-balanced deployment across multiple servers. Pitney Bowes reports that most enterprise customers process large, multi-million record data in few hours. Fault-tolerance is not inbuilt within the product. However Pitney Bowes reports that it can work well with the customer's existing fault tolerance infrastructure.

**Product Emphasis**

Pitney Bowes CDQP intends to become the solution of choice for maintaining the quality of customer data. In Butler Group's opinion, the solution suite focuses on two key aspects: 1) a high degree of integration with existing sets of data and data operations tools: 2) exclusive domain focus on customer data. In Butler Group's opinion, CDQP compares well with peers, and the company's history makes them particularly suitable for customer data integration projects.

## DEPLOYMENT

Pitney Bowes reports that implementation time can range from one week for simple deployments to four to six weeks for complex integration needs, business rule definitions, and full system testing. Once the solution is implemented, a data administrator would be required to apply data subscription updates on a monthly or quarterly basis; a system administrator would be required for product upgrades and patch updates; and business users for unit and system testing during and after upgrades.

Pitney Bowes maintains training facilities across North America, EMEA and APAC and provides on-premises training as well. Maintenance is provided through a tiered customer support process, and ranges from Web access via a support portal up to 24x7 on-call service. Professional services can be dispatched to a customer site or work remotely to help/create specific, complex flows.

The solution is available on Windows 2000, XP, and 2003; Solaris 8, 9, and 10; HP-UX 11.11, 11.23 (PA-RISC), and 11.23i (Itanium); AIX 5.2 and 5.3; Red Hat Linux 3 and 4; and Suse 9.

Risks related to CDQP deployment are similar to those associated with most enterprise software implementation projects, with a crucial variance in degree. Data integration and governance initiatives require significant executive sponsorship and inter-department collaboration, and projects are likely to suffer without an appropriate project governance mechanism.

## PRODUCT STRATEGY

The target market for the product is both vertical and horizontal. It includes B2C organisations (managing and interacting with more than 100,000 customers). Key vertical markets and sectors include financial services, Insurance, retail/restaurant/real estate, Telco and utilities, public sector, and emerging markets. In terms of company size, Pitney Bowes targets companies with over US\$10M in revenues.

Pitney Bowes reports that time to ROI is usually one year. ROI accrues from customer satisfaction and improved operational processes, which are realised through rapid and accurate data capture, single-customer view, and customer acquisitions made possible through data enrichment as part of marketing campaigns.

The product is sold primarily by Pitney Bowes' direct sales force. However, Pitney Bowes also uses channels like partners, OEM, and Value-added Resellers (VARs).

The key business partnerships and alliances that support this product are: Accenture, EDS, HCL, Headstrong, Infosys, Evaxys UK, Alicornia South Africa, BIT Mexico, and SOA Mexico. The key technology partnerships include: Siperian, Initiate Systems, SAP, Siebel, Oracle, Amdocs, ADP-Taxware, Microsoft, Silver Creek Systems, Composite Software, Business Data Quality UK, Convergys, CGI-AMS, and Accenture.

The company offers perpetual and term-based licences for on-premise deployments. The fees structure includes licence, maintenance, and subscription (annual) fees. The solution is also provided as a Software-as-a-Service model through hosted offering and licensing based on per-click usage that can be purchased upfront as an option. The company offers professional services for fast-start programs that may range from two to five days, to more comprehensive consulting support that may range from two to eight weeks. Services are normally 5% to 20% of the overall cost. The cost of annual maintenance and support ranges from 18% to 20%. This includes Technical Support and Product Upgrades. Optional 24x7 support is available as a premium add-on.

The upcoming November release will provide an Open Parser that will provide more control to users of the software to define parsing grammar to create structured metadata for any data domain with culture sensitive rules and local language constructs. This, along with their continued expansion of international geocoding capabilities expands the platform to provide support for more global opportunities and deployments. The release will also provide capabilities to “pull” remote services into local deployments giving companies access to value-added services both internal and external to the organisation. Certified integrations with mySAP CRM 2007 and Siebel Business and Industry Applications v.8 are also part of the release.

The future development planned for the product includes enhancement of Data Governance capabilities with Match Update Editor; tighter integration with profiling and monitoring tools; expansion of international geocoding support; support for promotion, versioning, and role-based security within the platform; addition of advanced spatial operations (routing and polygon-in-polygon); dynamic mapping for visualisation and analytics; demographics; and addition of more out-of-the-box templates.

Pitney Bowes’ CDQP strategy is to continue to focus organic product development on the customer data domain, tying the solution more closely with its location intelligence capabilities, and to partner with companies like Silver Creek Software to expand its offering outside the customer domain into areas such as Product Data Quality. In the area of Master Data Management (MDM), Pitney Bowes’ current strategy is to emphasise its offering and value in Data Governance/Quality/Enrichment as part of a larger MDM ecosystem, and as such it has partnered with MDM Hub vendors like Siperian, Initiate, Oracle, and SAP. Butler Group believes that given CDQP’s development pipeline and an installed base of 7,000 customers, Pitney Bowes is well placed to benefit from growing CDI requirements.

## COMPANY PROFILE

Pitney Bowes is a global organisation with headquarters in Stamford, CT. Pitney Bowes Business Insight (PBBI) has major offices in Lanham, MD, and Troy, NY in the US; Toronto, Canada; London, Watford, and Windsor, UK; Munich, Germany; Paris, France; Milan, Italy; Singapore; Tokyo, Japan; Noida, India; and Sydney and Melbourne in Australia.

Pitney Bowes Business Insight (PBBI) is a division of Pitney Bowes Inc., which is traded on NYSE under the symbol PBI. The division was formed in September 2007, by the unification of two industry leaders: Group 1 Software and MapInfo. PBBI’s solutions are available in multiple languages in 60 countries.

Pitney Bowes has more than 36,000 employees across the globe, out of which more than 1,700 are in the PBBI division. PBBI has more than 7,000 customers across the globe.

Currently more than 200 customers are using the Customer Data Quality Platform solution. The extent of implementation of the solution in these organisations varies from global in some cases to departmental in others. It can support real-time transactions of 1,400-1,800/sec in an enterprise deployment; 50-100 million records are processed in batch per month.

<b>Table 1: Financial Details</b>			
Year ending December 31	<b>2007</b>	<b>2006</b>	<b>2005</b>
<b>Revenue (US\$ Billion)</b>	6.12	5.73	5.37
<b>Change on Previous Year (%)</b>	6.8	6.7	-
<b>Total Net Income/(Loss) (US\$ Billion)</b>	1.66	0.47	2.19
Source: Pitney Bowes			<b>DATAMONITOR</b>

## SUMMARY

The data quality management solutions market place is an interesting place to watch. The market is populated with vendors from diverse backgrounds which leads to variance in add-on capabilities and variance in process focus, vendors focusing on specific data types, and vendors that are domain agnostic. The market includes behemoths such as IBM as well. Pitney Bowes' well documented success in all things related to location intelligence gives the company a sizeable advantage in customer data quality management. The benefit that a domain-independent data quality management solution would bring to an enterprise-wide data governance initiative is appreciated. However, Butler Group believes that feature-by-feature and architecture-wise the solution compares well with the best in the market, and the company's capabilities in related areas align the solution really well with customer data integration requirements.

Table 2: Contact Details	
<p><b>Pitney Bowes Business Insight</b>                      ATTN: Clarence Hemphield                      4200 Parliament PI                      Lanham, MD 20706                      USA                      Tel: +1 301.918.0747                      www.pbbusinessinsight.com</p>	<p><b>Pitney Bowes Business Insight</b>                      Minton Place                      Victoria Street                      Windsor, Berkshire, England, SL4 1EG                      UK                      Tel: +44-1753 848200                      Fax: +44-1753 621140                      www.pbbusinessinsight.com</p>
Source: Pitney Bowes	<b>DATAMONITOR</b>

**Headquarters**

Shirethorn House,  
 37/43 Prospect Street,  
 Kingston upon Hull,  
 HU2 8PX, UK  
 Tel: +44 (0)1482 586149  
 Fax: +44 (0)1482 323577

**Butler Direct Pty Ltd.**

Level 46, Citigroup Building,  
 2 Park Street, Sydney,  
 NSW, 2000,  
 Australia  
 Tel: + 61 (02) 8705 6960  
 Fax: + 61 (02) 8705 6961

**Butler Group**

245 Fifth Avenue,  
 4th Floor, New York,  
 NY 10016,  
 USA  
 Tel: +1 212 652 5302  
 Fax: +1 212 202 4684

**Important Notice**

This report contains data and information up-to-date and correct to the best of our knowledge at the time of preparation. The data and information comes from a variety of sources outside our direct control, therefore Butler Direct Limited cannot give any guarantees relating to the content of this report. Ultimate responsibility for all interpretations of, and use of, data, information and commentary in this report remains with you. Butler Direct Limited will not be liable for any interpretations or decisions made by you.

For more information on Butler Group's Subscription Services please contact one of the local offices above.

