Pitney Bowes
GROUP 1 SOFTWARE

Data Profiling: Underpinning Quality Data Management

White Paper

Data Profiling: Underpinning Data Quality Management

Overview

The current economic climate forces all businesses to compete more effectively. The ever increasing pressure to reduce margins in recent years has forced organizations to look inwardly at solutions based on cutting costs and maximizing customer profitability. These drivers result in a reliance on IT systems supporting the likes of: Enterprise Resource Planning, Customer Relationship Management, Supply Chain Management, Manufacturing, Stock Control, Logistics, and Business Intelligence Solutions, to name but a few.

However, all of these solutions only deliver value if the data they depend on is accurate, complete, and consistent. This information feeds and supports everything from the simplest business process to the highest level strategic decision. Hence, the quality of an organization's data assets directly impacts on the bottom line.

A recent survey commissioned by the Data Warehousing Institute estimates that poor data quality costs US businesses alone more than $600 billion a year. In our experience, the greatest losses are derived from:

- Wasted investment into new systems that do not deliver a return or concrete benefits
- Excessive overheads due to inflexible processes reliant on data that is not "fit for purpose"
- Lost revenue caused by poor quality customer data
- Flawed strategic business decisions based on inaccurate or incomplete data

The need to analyze data is at the foundation of every effective data management strategy since the dawn of modern information systems. Whether auditing your data assets to assess quality, ensure regulatory compliance, gain a better understanding of your information needs, or to embark upon new systems implementations, Data Profiling delivers a deeper and broader insight in a fraction of the time required by traditional approaches to data analysis.

Data Analysis Versus Data Profiling

Traditional data analysis techniques are simply unable to cope with the scale of today's data management challenges, which inevitably stretch the analysis resources available to the point where radical reductions of scope become a necessity. These factors contribute to the near certainty of missing completely, or managing incorrectly, critical data quality problems.

More often than not, data quality issues remain hidden until they surface at a time when they have the most negative impact on business performance. In a project context, this frequently occurs during loading, final testing, or go-live situations where the costs of retrospectively fixing the issues have risen exponentially (sometimes to the point where all the business value of the project is negated). It is impossible to guess in advance where these problems exist. Only comprehensive and regular data audits identify all possible dangers before they adversely affect the business or become unmanageable.

> *"Unfortunately, conventional methods for analyzing real data take a great deal of time, involve only small samples of the data and fail to deliver a complete understanding of the source data. Manual or semi-automated processing techniques cannot possibly compare the thousands of attributes and millions of values necessary to uncover the relationships. The answer is a new category of software called data profiling [...] which offers a fast, accurate and automated way to understand your data. It enables a small, focused team of technical and business users to quickly perform the highly complex tasks necessary to achieve a thorough understanding of source [or target] data. This level of understanding cannot be achieved through conventional approaches."*
>
> Craig Olson, Data Management Review, March 2000

Data profiling technology vastly improves the scope and depth of data analysis in three key ways:

- Automation of traditional analysis techniques – it is not uncommon to see analysis time cut by 90% while still providing a better understanding.
- Ability to apply a brute force approach to analyzing data – analysts are no longer limited to working just with sample data.

Terabytes of data are profiled effectively and completely. Sometimes the smallest anomalies have the greatest impact.

- Assessment of rules that govern data which cannot easily be discerned via manual coding and inspection – pattern generation, dependency testing, join analysis.

Data Profiling is the crucial first step to be undertaken at the start of any data-driven initiative, whether a new data warehouse, system migration, data integration project, or the implementation of a corporate data quality strategy. Without the level of insight Data Profiling provides, the risk of hitting data quality issues remain unacceptably high for modern businesses.

The Problems Data Profiling Addresses

The old adage, "garbage in, garbage out" applies more today than ever before as data-centric systems support every aspect of running a business. If data is of a poor quality, or managed in structures that cannot be integrated to meet the needs of the enterprise, business processes inevitably suffer, as will decision-making and profitability.

Data quality issues are endemic in most organisations – duplication, incompleteness, inconsistency, and countless other potential anomalies all play a part in undermining operational efficiency. Furthermore, any gaps in one's understanding of the business rules that govern that data result in new processes and systems being rolled out that fail to meet the ever-changing needs of the organization, thus compounding the existing problems, compromising trust in the data and increasing risk.

Without a thorough understanding of data quality issues, it is almost inevitable that development costs will spiral and projects will overrun, or even fail outright. A clear, up-front picture of all the potential issues is essential to plan projects effectively. Data cleansing and transformation requirements must be understood before timescales and costs are finalized, not after. Traditional approaches to data analysis simply cannot answer all the questions that need to be asked. Especially when it is not clear what those questions are in the first place.

It is not always easy to quantify the exact cost of having projects and business processes undermined by data quality issues, but it is accepted that the impact on profitability across an enterprise. Most disturbing of all is that many organizations have few metrics in place to assess the extent to which poor data quality affects the bottom line, whether that be from lost revenue, excessive overheads, or unsound business decisions based on misleading business intelligence. Ultimately, the success or failure of a business depends on the quality of its data assets.

Although often hidden within departmental budgets, it is easier to value the ongoing cost of "fire-fighting" systems exhibiting data quality issues. These costs break down principally to the associated requirements of:

- Identifying data quality issues through thorough data analysis
- Managing the priorities, statuses, and allocations of remedial activities and resources
- Implementing, validating, and sustaining data quality improvements

Even taken on its own, the ability to reduce or eliminate these overheads almost always justifies the adoption of Data Profiling technology. But, the consequent benefits across the enterprise are often many times greater.

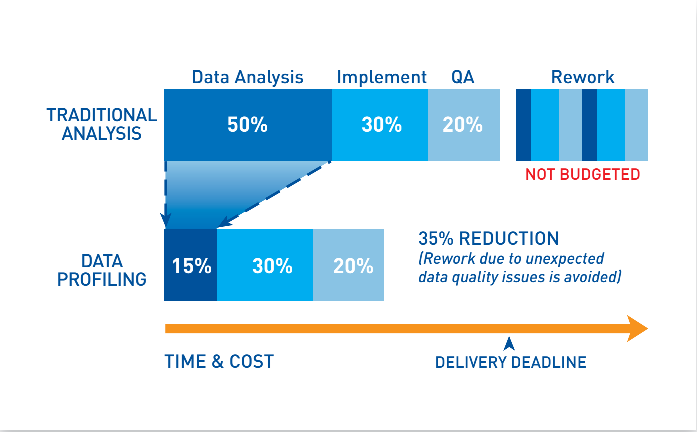


Figure 1: Typical Project-based Scenario

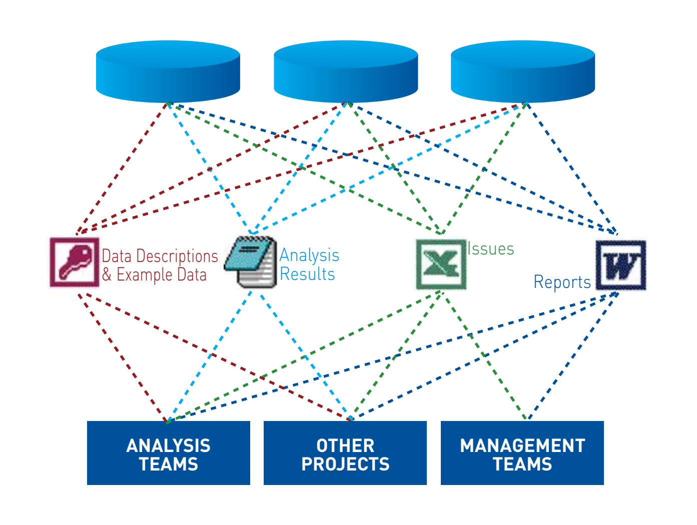


Figure 2: Traditional Analysis Approach

- >> No centralization of information
- >> Tools applied on 'ad hoc' basis
- >> High data management overheads
- >> Increased effort and greater risk

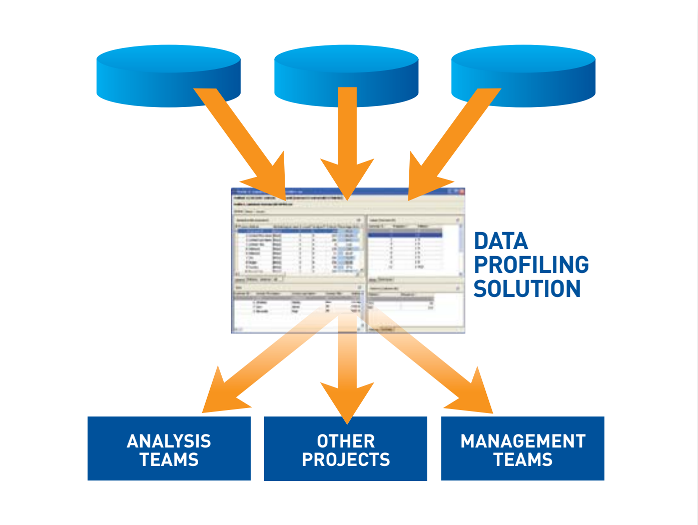


Figure 3: Data Profiling Led Approach

- >> Single point of reference
- >> Designed for analysis activities
- >> Reduced management overheads
- >> Increased confidence and lower risk

Implementing a Data Profiling led approach, as a first step in all data-driven projects, radically improves the performance of activities reliant on data management and significantly reduce risks, costs, and timescales.

When used, in tandem, to support a strategic data quality initiative, Data Profiling identifies gaps early to ensure that high performance is sustained across the organization.

How Does Data Profiling Promote Better Data Quality?

Delivering better data quality relies first and foremost on understanding the data you manage and the rules that govern that data. Without this knowledge, no effective data management plan can be formulated. Profiling data provides both the framework and roadmap to improved data quality, smoother running systems, more efficient business processes, and ultimately, the performance of the enterprise itself.

Compared to manual analysis techniques, Data Profiling technology significantly improves the organization's ability to meet the challenge of managing data quality.

Core Features of Data Profiling Technology

A complete Data Profiling solution delivers three-dimensional analysis rather than being limited largely to what is achievable through manual techniques, namely Column Profiling. Ideally, the solution should also support a collaborative approach that centralizes analytical effort, metadata management for structures and rules, issue tracking workflow, process management, documentation, and reporting.

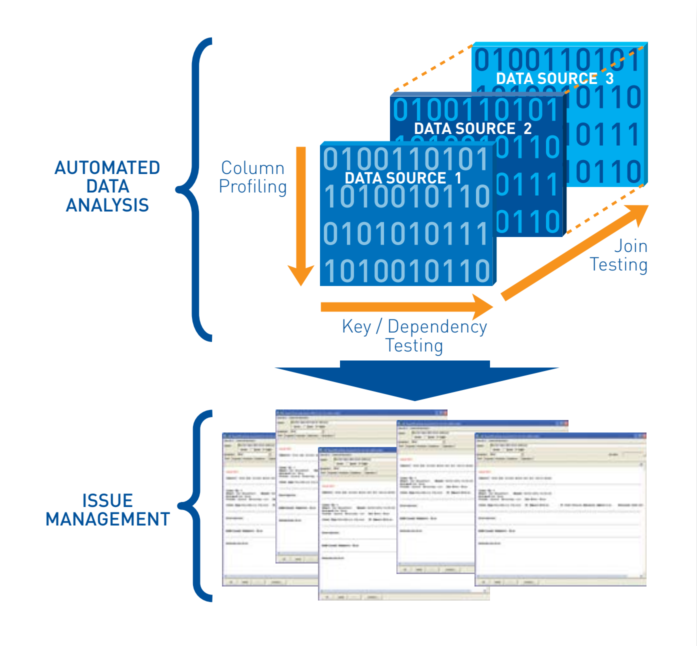
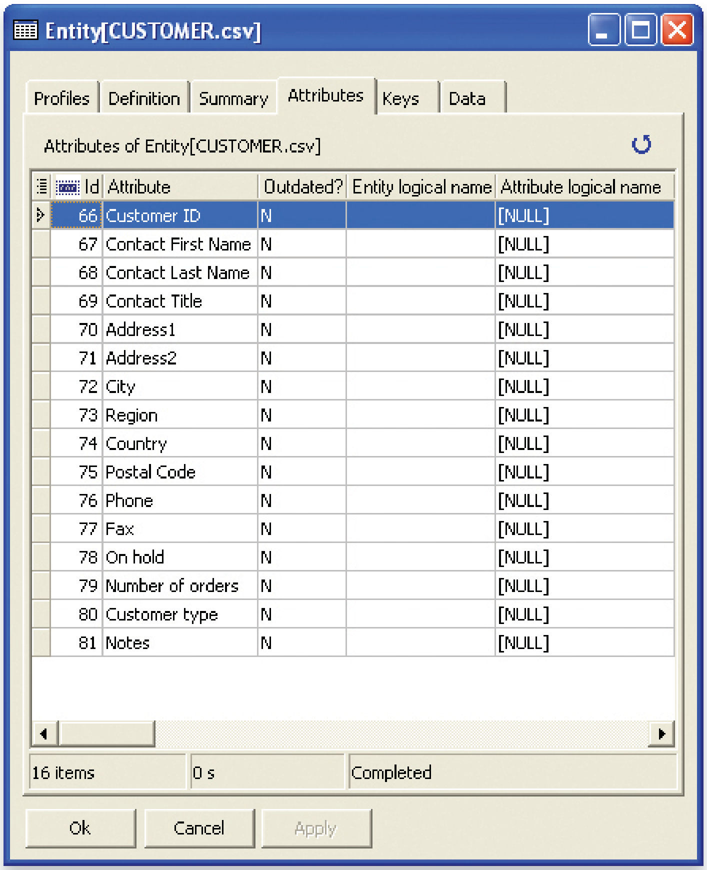


Figure 4: Core Features of Data Profiling Technology

Metadata Management

Managing metadata (i.e. the library of definitions describing the countless data structures and rules) is a non-trivial task. More often than not, this crucial information is both not accessible to those who need it, and out of date. Typically, these definitions are captured only when systems are implemented and stored in text documents that have no physical links to the systems themselves. Dynamically changing business needs lead to constant evolution, not only in the way existing structures are used and interpreted, but also in the behaviour of the data itself. New structures are added and some existing structures become redundant. Similar or related information is often stored in several parts of the organization. Maintaining accurate cross-references is pivotal to ensuring synchronization and consistency.

The advantage of Data Profiling technology is that it does not rely on existing system documentation or pools of expertise hidden in the business. Instead, insight into the structure and rules governing data assets is derived from the data directly, thus avoiding incorrect assumptions and overlooked issues.



Attributes of Entity[CUSTOMER.csv]

Id	Attribute	Outdated?	Entity logical name	Attribute logical name
66	Customer ID	N		[NULL]
67	Contact First Name	N		[NULL]
68	Contact Last Name	N		[NULL]
69	Contact Title	N		[NULL]
70	Address1	N		[NULL]
71	Address2	N		[NULL]
72	City	N		[NULL]
73	Region	N		[NULL]
74	Country	N		[NULL]
75	Postal Code	N		[NULL]
76	Phone	N		[NULL]
77	Fax	N		[NULL]
78	On hold	N		[NULL]
79	Number of orders	N		[NULL]
80	Customer type	N		[NULL]
81	Notes	N		[NULL]

16 items | 0 s | Completed

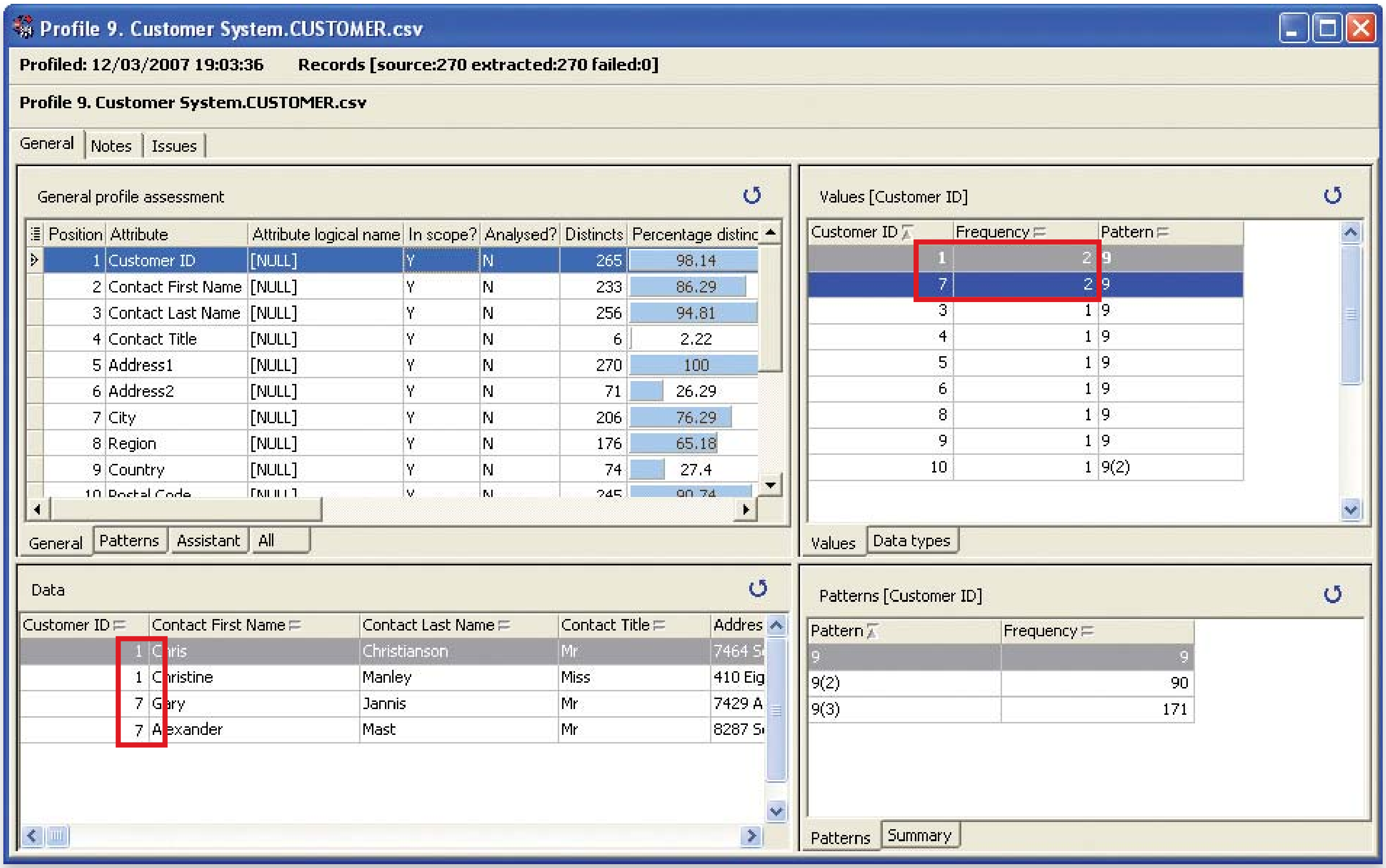
Figure 5: Metadata Management

Comprehensive Data Profiling solutions will provide a framework for capturing and maintaining an accurate Data Dictionary, driven by the data resources themselves, across all source platforms.

Ideally, they should additionally provide the option of managing snapshots of data sources (along with analysis results) offline from operational systems. This allows the review of data quality histories can be reviewed over time to identify trends and spot unexpected anomalies as soon as they develop.

Column Profiling

Column profiling provides the first cut in understanding data. This



Profile 9. Customer System.CUSTOMER.csv

Profiled: 12/03/2007 19:03:36 Records [source:270 extracted:270 failed:0]

Profile 9. Customer System.CUSTOMER.csv

General profile assessment

Position	Attribute	Attribute logical name	In scope?	Analysed?	Distincts	Percentage distinc
1	Customer ID	[NULL]	Y	N	265	98.14
2	Contact First Name	[NULL]	Y	N	233	86.29
3	Contact Last Name	[NULL]	Y	N	256	94.81
4	Contact Title	[NULL]	Y	N	6	2.22
5	Address1	[NULL]	Y	N	270	100
6	Address2	[NULL]	Y	N	71	26.29
7	City	[NULL]	Y	N	206	76.29
8	Region	[NULL]	Y	N	176	65.18
9	Country	[NULL]	Y	N	74	27.4
10	Postal Code	[NULL]	Y	N	245	90.74

Values [Customer ID]

Customer ID	Frequency	Pattern
1	2	9
7	2	9
3	1	9
4	1	9
5	1	9
6	1	9
8	1	9
9	1	9
10	1	9(2)

Data

Customer ID	Contact First Name	Contact Last Name	Contact Title	Addres
1	Chris	Christianson	Mr	7464 S
1	Christine	Manley	Miss	410 Eig
7	Gary	Jannis	Mr	7429 A
7	Alexander	Mast	Mr	8287 S

Patterns [Customer ID]

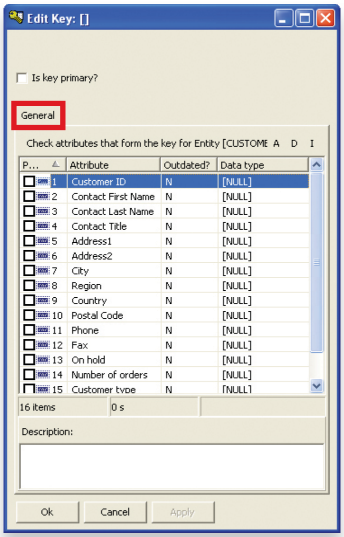
Pattern	Frequency
9	9
9(2)	90
9(3)	171

Figure 6: Column Profiling

is the area where traditional manual data analysis techniques focus most of their attention, looking at each data attribute (column) in turn to evaluate basic features such as dominant data type, percentage population, uniqueness, value ranges, and field lengths.

However, unlike manual analysis, Data Profiling provides several additional capabilities that cannot easily be achieved using traditional, hand-coded approaches.

- The ability to process the entire source rather than a subset of rows or columns limited by how much time is available to the team.
- All data features across the source can be summarized and reviewed rather than just focusing on the issues that are already known.
- Generation and analysis of patterns in data. For instance, many items such as postal codes and customer identifiers conform to a defined set of standard alphanumeric structures. Pattern analysis enables immediate identification of non-standard values.
- Most Data Profiling tools automatically generate value-frequency lists allowing analysts to quickly review the range and spread of data values as well as identifying unexpected duplicates.



☐ Is key primary?

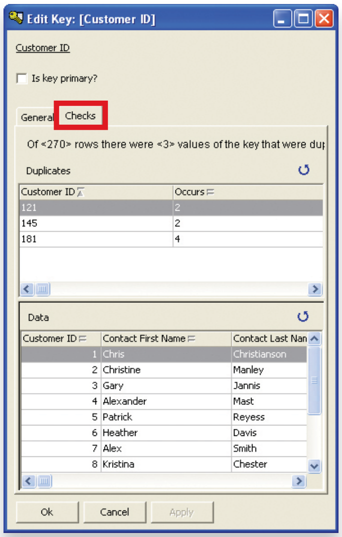
Check attributes that form the key for Entity [CUSTOME A D I

P...	Attribute	Outdated?	Data type
☐ 1	Customer ID	N	[NULL]
☐ 2	Contact First Name	N	[NULL]
☐ 3	Contact Last Name	N	[NULL]
☐ 4	Contact Title	N	[NULL]
☐ 5	Address1	N	[NULL]
☐ 6	Address2	N	[NULL]
☐ 7	City	N	[NULL]
☐ 8	Region	N	[NULL]
☐ 9	Country	N	[NULL]
☐ 10	Postal Code	N	[NULL]
☐ 11	Phone	N	[NULL]
☐ 12	Fax	N	[NULL]
☐ 13	On hold	N	[NULL]
☐ 14	Number of orders	N	[NULL]
☐ 15	Customer type	N	[NULL]

16 items 0 s

Description:

Figure 7: Key Testing



Customer ID

☐ Is key primary?

Of <270> rows there were <3> values of the key that were du

Duplicates

Customer ID	Occurs
121	2
145	2
181	4

Data

Customer ID	Contact First Name	Contact Last Nan
1	Chris	Christianson
2	Christine	Manley
3	Gary	Jannis
4	Alexander	Mast
5	Patrick	Reyess
6	Heather	Davis
7	Alex	Smith
8	Kristina	Chester

Figure 8: Key Testing

>> Instant “drill-down” access to relevant records based on identified features (a value or pattern) that require further analysis.

Figure 6 illustrates a column profile view where duplicate Customer ID numbers are identified. The bottom-left pane contains the drill-down to affected records.

Many products that claim to be Data Profiling solutions only provide column profiling capabilities. To achieve a complete understanding of your data, it is essential to look beyond the behavior of attributes in isolation so that relationships between attributes within and across tables/systems are assessed and tested for anomalies.

Key Testing

Key testing allows the analyst to verify the uniqueness of primary and candidate keys within an entity (table). Since keys uniquely identify the data in other non-key fields, it is crucial that they perform the intended role. Key testing provides an explicit tool for assessing more complex situations – in particular, where keys are defined using combinations of more than one attribute.

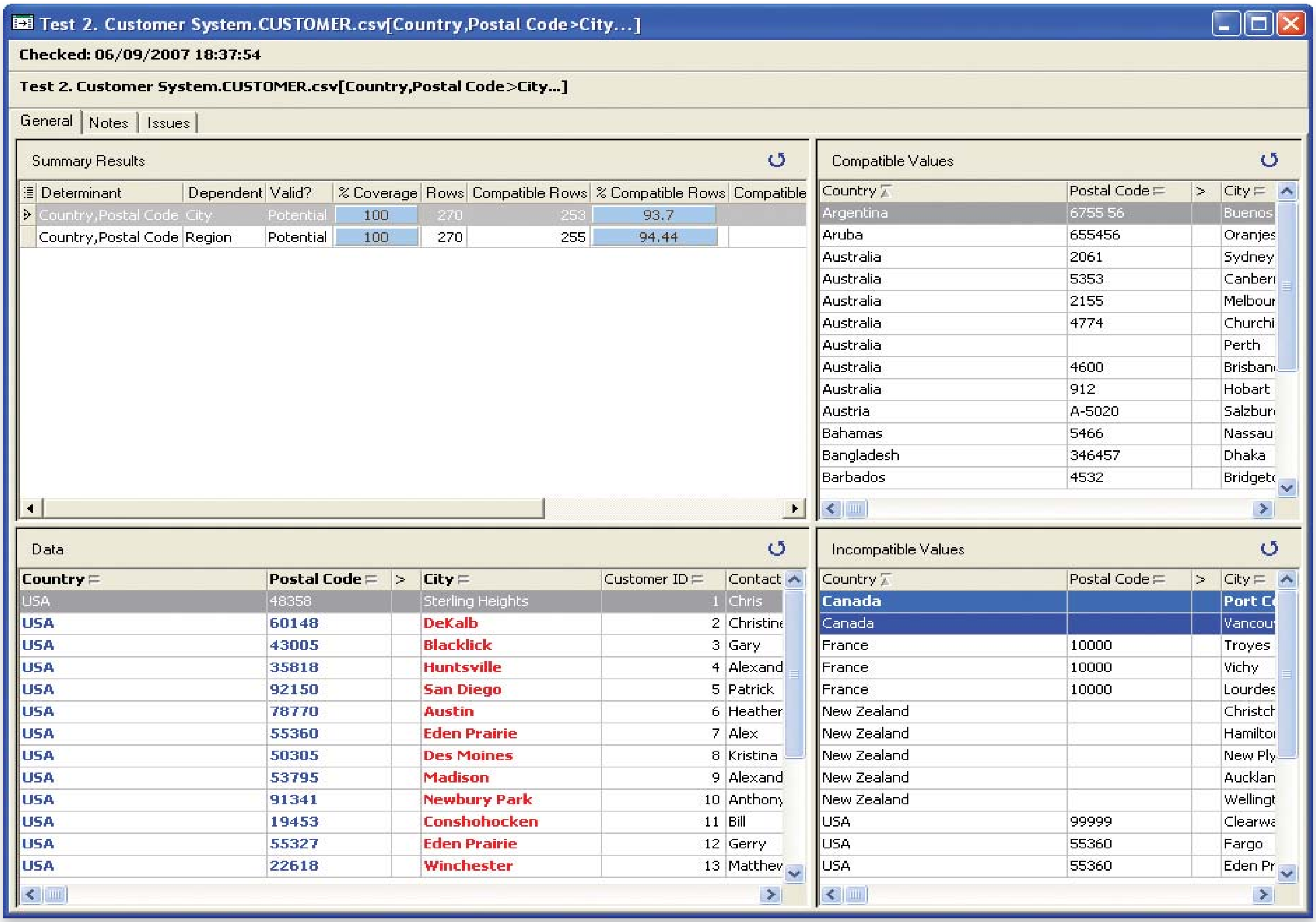
In Figures 7 and 8, we show an alternative key combination being tested. Again, note how Data Profiling allows the user to interactively define and test the key and drill down to exceptions in a fraction of the time it would take to script an equivalent query and run it against the database.

Dependency Testing

Like key testing, dependency testing focuses on the relationships between data values across attributes (columns) in a single entity (table). However, here the emphasis is not on uniqueness, but consistency. For instance, if we believe that a combination of two field values [a,b] should always specifically match values in three other fields [x,y,z], we can test this correlation regardless of rows where the same values are duplicated. We are only interested in exceptions such as [a,b] determines both [x,y,z] and [x,y,y].

Dependency testing has particular relevance when restructuring data. The majority of older systems relied on merging tables (de-normalization) to enhance performance. This results in the values in some columns not being directly dependent on the primary key, but on other column(s) in the table which, in turn, are dependent on the key (transitive dependencies). Although this technique offers performance benefits, it also leads to anomalies creeping into the data. These anomalies cause problems if the data is subsequently moved into more optimised structures (a Data Warehouse or new ERP system), or if the data is just used in new ways and assumptions are being made about its consistency.

Figure 9 shows the result of a dependency test where the combination of specific Country and Postal Code values are always expected to return consistent City and Region values. For instance, a combination of Country UK and Postcode MK1 should always occur together with the City Milton Keynes. The example shows that City and Region values correlate only 97.04% and 97.78%, respectively. Both relationships are thus classified as “Potential” and all compatible and incompatible value combinations are shown in the panes on the right. The highlighted rows show a drill-down to a specific pair of inconsistent rows. It is clear from looking at this example that the primary cause of the inconsistency is that there are missing Postal Codes leading to ambiguity in the dependency being tested. One might chose to retest this relationship after the issue of missing Postal Codes is addressed.



Test 2. Customer System.CUSTOMER.csv[Country,Postal Code>City...]

Checked: 06/09/2007 18:37:54

Test 2. Customer System.CUSTOMER.csv[Country,Postal Code>City...]

Summary Results

Determinant	Dependent	Valid?	% Coverage	Rows	Compatible Rows	% Compatible Rows	Compatible
Country,Postal Code	City	Potential	100	270	253	93.7	
Country,Postal Code	Region	Potential	100	270	255	94.44	

Compatible Values

Country	Postal Code	>	City
Argentina	6755 56		Buenos
Aruba	655456		Oranjes
Australia	2061		Sydney
Australia	5353		Canberr
Australia	2155		Melbour
Australia	4774		Churchi
Australia			Perth
Australia	4600		Brisban
Australia	912		Hobart
Austria	A-5020		Salzbur
Bahamas	5466		Nassau
Bangladesh	346457		Dhaka
Barbados	4532		Bridget

Data

Country	Postal Code	>	City	Customer ID	Contact
USA	48358		Sterling Heights	1	Chris
USA	60148		DeKalb	2	Christine
USA	43005		Blacklick	3	Gary
USA	35818		Huntsville	4	Alexand
USA	92150		San Diego	5	Patrick
USA	78770		Austin	6	Heather
USA	55360		Eden Prairie	7	Alex
USA	50305		Des Moines	8	Kristina
USA	53795		Madison	9	Alexand
USA	91341		Newbury Park	10	Anthony
USA	19453		Conshohocken	11	Bill
USA	55327		Eden Prairie	12	Gerry
USA	22618		Winchester	13	Matthew

Incompatible Values

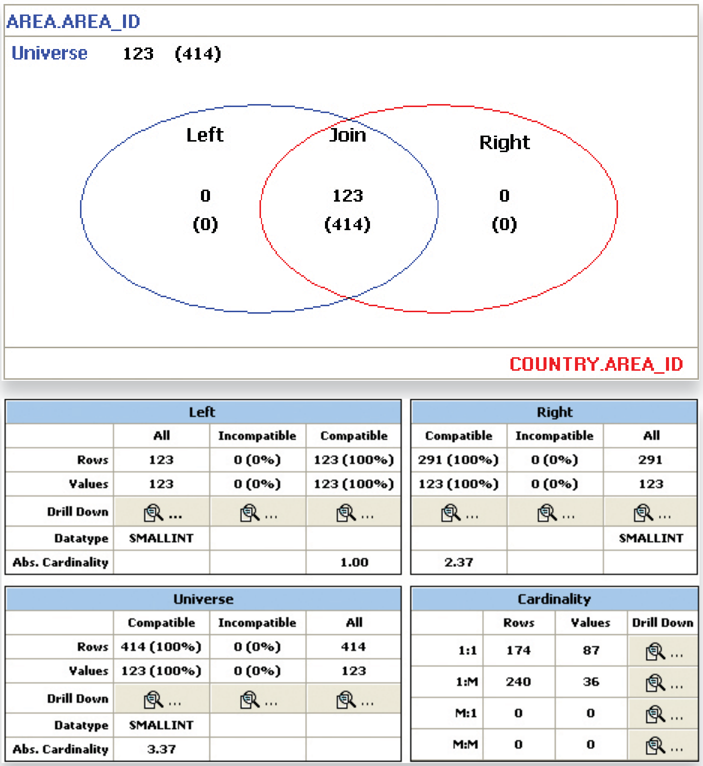
Country	Postal Code	>	City
Canada			Port C
Canada			Vancou
France	10000		Troyes
France	10000		Vichy
France	10000		Lourdes
New Zealand			Christch
New Zealand			Hamilto
New Zealand			New Ply
New Zealand			Auckland
New Zealand			Wellingt
USA	99999		Clearwa
USA	55360		Fargo
USA	55360		Eden Pr

Figure 9: Dependeny Testing

Join Testing

Join testing is a vital step in any project relying on the ability to integrate data from separate sources. Consider a scenario where a data warehouse is envisaged to create a single view of customer data. The requirements may dictate that data is drawn from previously unconnected systems managing customer relations, accounts, ordering, and deliveries. Any assumptions about the common data (such as customer and account codes) that are used to merge the sources must be thoroughly tested as part of assessing the feasibility of the project. Figures 10 and 11 provide more detail.

Figure 10 shows a clean result from a join test where there is a complete overlap of AREA_ID values between a Sales Area table containing 123 records (all with unique values) and a Country lookup table containing 291 records (containing the same 123 unique values). The fact that there are duplicates in the Country

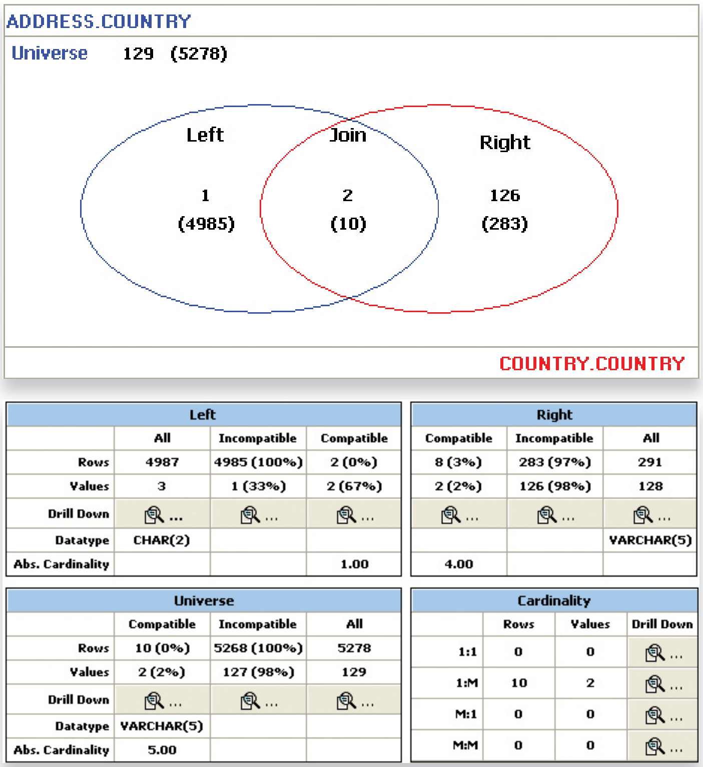


	Left			Right		
	All	Incompatible	Compatible	Compatible	Incompatible	All
Rows	123	0 (0%)	123 (100%)	291 (100%)	0 (0%)	291
Values	123	0 (0%)	123 (100%)	123 (100%)	0 (0%)	123
Drill Down	🔍 ...	🔍 ...	🔍 ...	🔍 ...	🔍 ...	🔍 ...
Datatype	SMALLINT					SMALLINT
Abs. Cardinality			1.00	2.37		

	Universe		
	Compatible	Incompatible	All
Rows	414 (100%)	0 (0%)	414
Values	123 (100%)	0 (0%)	123
Drill Down	🔍 ...	🔍 ...	🔍 ...
Datatype	SMALLINT		
Abs. Cardinality	3.37		

	Cardinality		
	Rows	Values	Drill Down
1:1	174	87	🔍 ...
1:M	240	36	🔍 ...
M:1	0	0	🔍 ...
M:M	0	0	🔍 ...

Figure 10: Clean Result from a Join Test



	Left			Right		
	All	Incompatible	Compatible	Compatible	Incompatible	All
Rows	4987	4985 (100%)	2 (0%)	8 (3%)	283 (97%)	291
Values	3	1 (33%)	2 (67%)	2 (2%)	126 (98%)	128
Drill Down	🔍 ...	🔍 ...	🔍 ...	🔍 ...	🔍 ...	🔍 ...
Datatype	CHAR(2)					VARCHAR(5)
Abs. Cardinality			1.00	4.00		

	Universe		
	Compatible	Incompatible	All
Rows	10 (0%)	5268 (100%)	5278
Values	2 (2%)	127 (98%)	129
Drill Down	🔍 ...	🔍 ...	🔍 ...
Datatype	VARCHAR(5)		
Abs. Cardinality	5.00		

	Cardinality		
	Rows	Values	Drill Down
1:1	0	0	🔍 ...
1:M	10	2	🔍 ...
M:1	0	0	🔍 ...
M:M	0	0	🔍 ...

Figure 11: Mismatch from a Join Test

lookup is not a concern as long as each Area matches a Country. The cardinalities in the Details table indicate a consistent relationship where records either match 1-to-1 or 1-to-Many (1:M overall). Note that, even with a full overlap of values in the Venn Diagram, any combination of values exhibiting both 1:M and M:1, or any M:M, relationships make integrating this data almost impossible. The reason is because it is not clear which Area record matched which Country record when alternatives exist on both sides.

Figure 11 shows a very different picture. In this case, there is a radical mismatch of Country values between an Address table and the same Country lookup we used above. Only 2 values are common (accounting for a mere 10 records between the tables).

We are not surprised to see that there are orphan Country codes if we already know that we only do business with a handful of countries in the lookup. However, there is an assumed business rule that all addresses we keep must match a country. This test has proved that all but two do not match. Using drill-downs we soon establish that the orphan Country value in the Address table is UK. Similarly, we uncover an unused orphan in the Country table of GB. Although these values have slightly different semantic meanings, we conclude that two systems were implemented using different standards and these values are assumed equivalent for the purposes of the project.

A prudent choice in light of this discovery is to standardize the values across the systems as a cleansing task and retest the join to ensure consistent cardinalities.

Note that well-implemented Data Profiling solutions automatically support join tests between attributes even if their names and datatypes do not match exactly.

Issue Management

Managing issue progress is a challenge in itself. In a single view of a customer data warehouse project run at a leading telecom provider in the UK, a team of data analysts identified 300 issues per week. As with most projects, issues are managed separately to the analysis process using tools like Excel or "home-grown" databases. In either case, tracking issues back to source data is practically impossible and managing the issue list becomes yet another overhead.

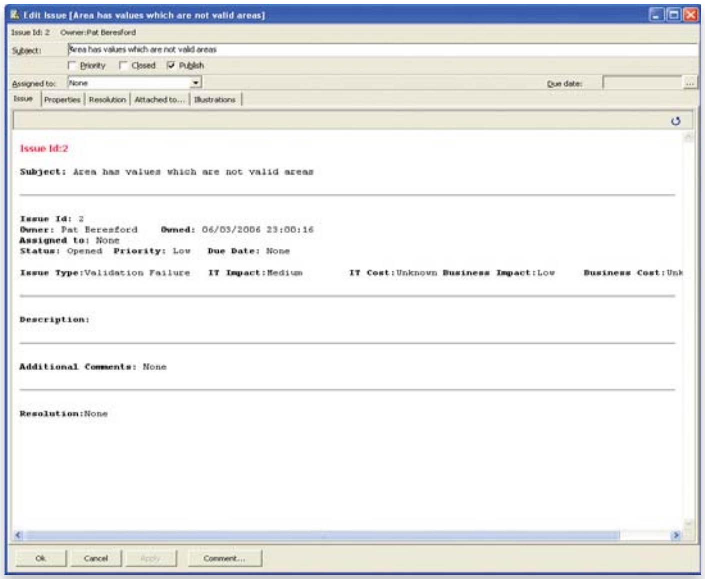
No Data Profiling technology is complete without integrated support for documenting and managing the lifecycle of data quality issues. Although it is often the case that cleansing and transformation tasks are carried out using other cherry-picked tools, it makes sense to have a central shared repository of all open and closed issues. It also makes sense to provide this functionality where the issues are identified so that they can be linked in the repository to the source structures and records that exhibited the problems.

Few Data Profiling offerings provide more than the option to attach simple notes to data structures and values. Fewer still support issue resolution workflow where issues can be attributed with details such as ownership, assignment, priority, status, categorisation, and their histories tracked for management purposes. Add to this the ability to report on issues by any combination of these features and one can provide complete visibility for issue management.

Deploying Data Profiling Technology

Data Profiling technology is readily integrated within existing data management functions. It supports, accelerates, and enhances most activities related to data analysis. Analysts achieve their goals to a greater depth in a fraction of the time normally required with such tasks.

In the context of implementing new systems and solutions, Data Profiling becomes part of the project toolkit along with other potentially necessary technologies such as Cleansing and ETL (Extract Transform Load) capabilities. But, due to the enhanced ability for auditing data sources relevant to the project, Data Profiling provides a much clearer view of existing quality issues at a time when decisions about downstream project activities are still being finalized. Therefore, more informed decisions can be made about the project timescales, approach, and which other technologies are required for successful project completion. In many cases, being able to perform a complete, rather than partial, data assessment forms a crucial element in qualifying the feasibility of the entire project.



Edit Issue [Area has values which are not valid areas]

Issue Id: 2 Owner:Pat Beresford

Subject: Area has values which are not valid areas

☐ Priority ☐ Closed ☑ Publish

Assigned to: None Due date:

Issue Id:2

Subject: Area has values which are not valid areas

Issue Id: 2
Owner: Pat Beresford Owned: 06/03/2006 23:00:16
Assigned to: None
Status: Opened Priority: Low Due Date: None

Issue Type:Validation Failure IT Impact:Medium IT Cost:Unknown Business Impact:Low Business Cost:Unk

Description:

Additional Comments: None

Resolution:None

Figure 12: Issue Management

Figure 13 illustrates how Data Profiling technology is applied within the data analysis and preparation phase of a typical project.

Beyond specific projects, Data Profiling also delivers significant benefits to any user engaged in assessing data on even an ad-hoc basis. Whether dealing with large or small volumes of data, profiling accelerates analytical work and provides a bridge for communication between technical, business and management members of a team.

Furthermore, as companies increasingly look to implement ongoing strategic data quality initiatives, Data Profiling technology offers existing corporate analysis and data management resources the required tools. These tools include the ability to analyze data within meaningful timeframes and the means by which to monitor progress and performance. Investing in data assets is only as valuable as the ability to sustain quality going forwards.

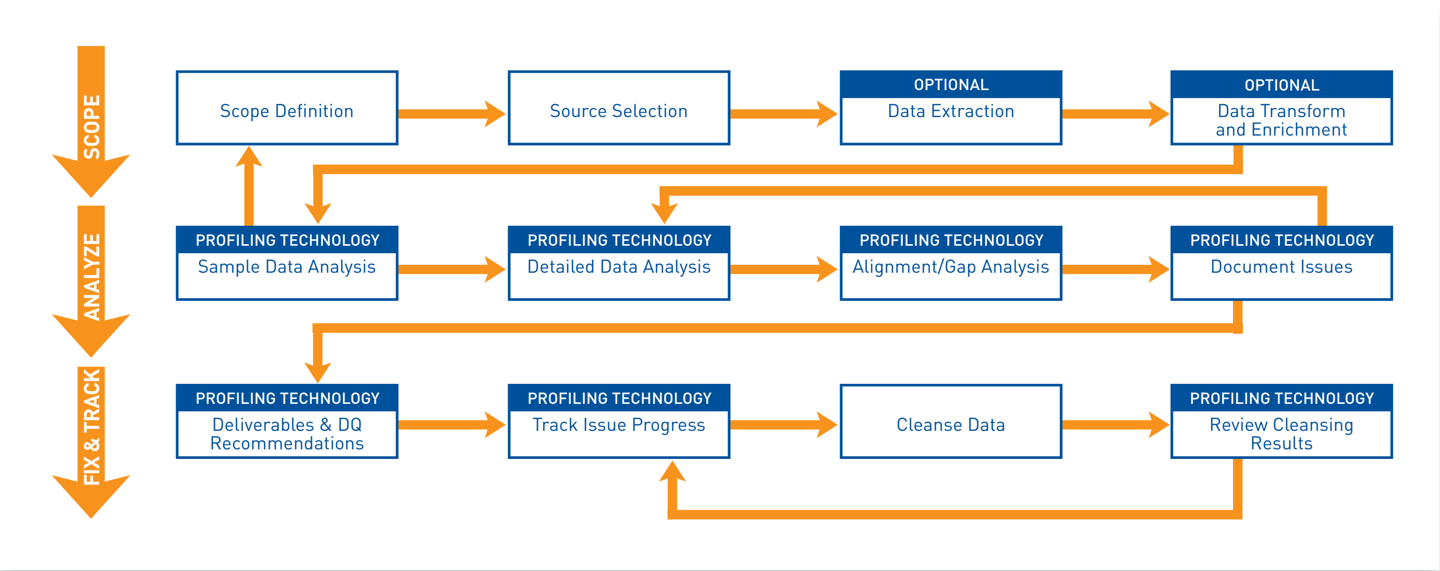


Figure 13: Applying Data Profiling Technology

Conclusion

Who Does Data Profiling Benefit, and How?

As a mature technology, it is amply demonstrated that a Data Profiling led approach delivers tangible value across the business when applied to the challenges of data analysis and quality management. Adoption is straightforward and the potential returns on investment very significant. At the enterprise level, its ability to raise and maintain the quality of corporate information promotes competitive advantages and cost cuts.

The following summarizes the direct benefits that are expected:

Data Analysts and Data Managers

>> Step improvements in analysis performance through automation – do more in less time

>> Significant increase to achievable breadth and depth of analysis scope

>> Far clearer understanding of data content and business rules

>> Facilitated communication between analysts, business users, and quality managers

Project Managers

>> Visibility of all data quality issues and their current statuses

>> Condensed and achievable delivery timescales

>> Reduced risk of project delays and budget over-runs due to unexpected data quality issues

System Owners

>> Sustainable data quality

>> Diminished operational costs

>> Fewer disruptions and less manual intervention

>> Ability to deliver a better value service

Data Owners/Stewards

- >> Framework for effective delivery of data quality strategy
- >> Ability to meet data quality responsibilities
- >> Greater confidence in data assets

Executive Management

- >> Effective business processes based on accurate, complete, and trustworthy information
- >> Better guarantee of return on investment from corporate systems and data assets
- >> Reliable information supports better strategic and tactical business decisions
- >> Increased profitability through improved efficiency and customer management

For more information about our products and services, please log onto our website at ***www.g1.com*** or call us today at ***888-413-6763***.

Pitney Bowes
GROUP 1 SOFTWARE

4200 Parliament Place, Suite 600
Lanham, MD 20706-1844
1-888-413-6763 • www.g1.com