Business Drivers for Data Quality in the Utilities Industry
ABSTRACT

While the traditional uses of data cleansing and address standardization within the utilities industry (such as postage cost savings) remain relevant, both market factors and changes are rapidly morphing the demand for information usability. The introduction of residential smart metering creates a drastic change in data collection from manual checks on a broad (e.g. monthly) scale to real-time communication of consumption on an almost continuous basis. At the same time, increased eco-awareness opens the door for more comprehensive customer statements, including detailed usage statistics as well as comparative reporting. And as energy utilities provide more detailed usage statistics to consumers, they gradually transform themselves from being providers of a commodity product to becoming a provider of information about energy.

And importantly, companies will need to have a more comprehensive view of their customers and their network, understand consumption and behavior models, develop creative product bundles and pricing models, as well as monitor many alternate data streams to help in anticipating (and rapidly responding to) outages, failures, and losses. A key aspect is the phenomenon Gartner refers to as “energy technology consumerization,” where energy consumers take active roles in managing their consumption as well as adopting renewable alternatives to energy generation. The increased percentage of renewable energy will make management of energy networks much more complex and will require more high quality information flowing in both directions to fuel the effective management of the network to address issues in real time.

In each of these scenarios, a successful business strategy leverages high quality information to degrees that go way beyond address standardization. This paper will look at emerging opportunities in the utilities industry yet see how they all still rely on high quality information. We then look at the types of capabilities that are necessary for data quality assurance, and provide some suggestions for consideration when putting a data quality plan in place.
The increased percentage of renewable energy will make management of energy networks much more complex and will require more high quality information.

Emerging Business Opportunities in the Utilities Industries

Utility companies are increasingly modernizing their approaches to taking advantage of their data to improve service, utilization, and customer experience. And as smart grid technologies are deployed, data volumes and reporting frequency are increasing, opening new opportunities for analyzing information to drive competitive advantage along dimensions of value such as improved customer experience, reducing costs, and increasing productivity. Some examples include:

- **Customer behavior analysis** – The desire to deliver highest-quality service while simultaneously seeking opportunities to develop new products, services, and bundles requires a better understanding of who the customers are and how they behave. The need is even more important for utilities in a competitive retail market. Customer behavior analysis looks at residential customer consumption patterns and sensitivities to better understand customer service expectations as well as model impact of price sensitive rate structure on customer demand.

- **Comparative usage reporting** – In reaction to increased consumer sensitivity to use of clean energy sources, energy consumption, utilities look to provide reporting that helps customers compare their consumption to similar customers in the same geographic region.

- **Asset optimization** – Location and management to more effectively track the lifecycle of hardware assets, average time between failures, and more effective stocking of parts for maintenance purposes and to help diagnose or predict transformer issues or monitor substation equipment maintenance and replacement needs.

- **Capacity monitoring, management, and allocation** – This includes modeling consumption to better manage the network and anticipate failures, losses, and outages, as well as monitoring usage and using historical information to dynamically allocate capacity across less congested parts of the network. This also helps in improving asset utilization through demand response and adjusting service and rates during peak times.

- **System modeling** – Using consumption and demand data for modeling of the loading on the transmission and distribution network to help in envisioning and evolving improvements in the grid infrastructure.

- **Energy technology consumerization** – Looking at consumer-deployed generation including solar panels or wind generation, understanding these customers and their expectations, and ultimately integrating their deployments and storage into the grid.

- **Customer experience management & behavior analysis** – This type of analysis integrates different sources of customer data to help analyze price sensitivity, support VIP Monitoring (identification of high-end or high-value customers to ensure the highest levels of customer service), and guide the bundling of commodities and services to better suit customer needs.

- **Managing incentives for consumer-generated energy** – Municipal, county, and state governments frequently offer incentives for consumers choosing to generate their own energy; managing the usage of privately-generated energy vs. provided energy becomes another area requiring improved reporting and analysis.

- **Business Impact Analysis** – Prediction of impacts of events or faults in the system to revenues, both from the infrastructure perspective with respect to pattern analysis that can predict equipment failure, as well as from the business perspective in seeking to optimize business processes to increase previously uncollected revenues, reduce payment delinquencies, and manage capital expenditures.
Traditional Opportunities Map to the Utilities Industry

Despite the excitement generated by the introduction of smart grid technologies for increased analytical capabilities, there remains more traditional types of reporting and analytics that are already being deployed in the utilities industry, or can be adopted from other, similar industries. In fact, as deregulation continues to expand the competitive market, increased attention to exploiting actionable information can improve these companies’ abilities to execute against their strategic plans, especially as more data sources (such as smart meter data or social media data) become available. These examples focus on facets of corporate value related to risk and regulatory compliance, improving the customer experience, and creating new opportunities for generating revenue (see chart below):

Best Practices in Data Quality

While the types of analyses presented in this paper may be intended to address different types of business challenges, all of these share two key characteristics: the need for comprehensive data about key concepts of the business such as, asset, commodity, revenue, or customer, and the requirement that this data be of high quality.

For example, customer behavior analysis, experience management, loyalty management, and upselling and/or product/service bundles all rely on accurate customer information. Incentive management, comparative reporting, and accurate compliance with billing statements and associated taxes require accurate data about the customers and their usage and consumption. Asset optimization, capacity management, and system modeling rely on high

<table>
<thead>
<tr>
<th>BILLING AND TAX MANAGEMENT</th>
<th>CUSTOMER PROFITABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government tax charges must be integrated into each customer’s statement. Residential customer locations are usually easily mapped to the proper municipal, county, and state jurisdictions. However, more complex infrastructures often span multiple jurisdictions (such as fences) or reside in “non-addressable” locations (such as mobile telecommunication towers), and this adds complexity to the generation of customer statements, requiring more sophisticated methods for analysis and reporting to ensure compliance with all jurisdictional tax requirements.</td>
<td>While customer retention is critical in competitive retail markets, the company may seek to assess the profitability of customer service, offer, and other aspects of customer churn models. For example, the costs associated with retention (based on premium offers or discounts) coupled with ongoing service costs may exceed the lifetime value for certain types of customers. This type of analysis will help to consider both managing service costs in a regulated market or limiting the spend for customer retention in a non-regulated market.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CUSTOMER LOYALTY</th>
<th>UPSELLING AND PRODUCT BUNDLING</th>
</tr>
</thead>
<tbody>
<tr>
<td>The introduction of competition in the energy industry will be coupled with the introduction of customer attrition—those who opt to select an alternate provider. Energy companies must have better algorithmic models for analyzing customer behavior, prediction of attrition, and use analytics and reporting to provide the right offers and incentives for building and maintaining customer loyalty.</td>
<td>There are emerging opportunities for creating new products and services and presenting them to the different types of customers. For example, identifying those customers whose consumption patterns are indicative of an owner of an electric vehicle may present opportunities for selling additional alternative services intended for the eco-conscious consumer. Another example is the installation of smart home-energy devices that allow the company to reduce in-home consumption during peak periods.</td>
</tr>
</tbody>
</table>
quality data about the network, hardware assets, and trucking fleets (especially for accurately stocking parts on repair trucks), as well as accurate geographic and location data.

Instituting Best Practices

As utility companies continue to broaden their offerings across different services and areas of their businesses, the demand for high quality data will only increase, and this is particularly true for those processes with customer touch points: customer profiling, loyalty management, and VIP monitoring, etc. To satisfy the need for assuring and maintaining high quality information, organizations will need to establish best practices in data quality management, such as:

- Data quality assessment, in which data sets are objectively reviewed using a statistical quantitative analysis. This helps data analysts pinpoint potentially flawed data such as outliers, missing values, data elements that do not conform to defined patterns or value domains, or are inconsistent with respect to other data elements.

- Data validation, in which data inconsistencies, missing values, or format variations are flagged as potential issues as records are created or imported, allowing the analysts to determine their criticality and take some corrective measures.

- Inspection and monitoring, in which defined rules are used for monitoring data quality and data audits are performed and reported to the proper data stewards within the organization.

These practices, which can be instituted as part of an operational data stewardship program, must enable the organization to incrementally adapt data quality management to legacy environments as well as map to newly-emerging technologies. For example, analyzing consumption behavior provides one view into customer behavior. However there are many potential sources of information about the customer, ranging from internal data from the call center, financial stability information from the credit and finance department, demographic data associated with residential locations, as well as data from third-party providers about individuals, locations, etc. Social media provide another channel for consumer information, such as interactions through sites such as Facebook or Twitter, etc. As more big data sources are acquired, the need for data quality increases, especially for customer experience management, such as parsing out relevant values from social media data and linking recognized identities to established customer profiles.

The Data Quality Toolkit

To support these best practices, organizations must arm themselves with the right tools for data quality management, such as:

- Parsing & Standardization
- Identity Resolution
- Merging of Duplicate Records
- Data Enhancement
- Data Cleansing
Parsing and Standardization
In many instances, data values are expected to conform to predefined sizes and formats, but slight variations in the data may lead to confusion. For example, state data elements may be represented using a variety of different formats, including character strings, digital codes, or character codes; some use punctuation or special characters. Humans may recognize the differences and similarities, but business applications may be less adept at determining whether these values are accurate or correct.

Parsing is a process used to describe formatting patterns and is used to determine whether a value conforms to a recognizable pattern as part of the assessment, matching and cleansing process. Parsing uses defined patterns, regular expressions, or grammars managed within a rules engine along with table lookups to distinguish between valid and invalid data values. When patterns are recognized, other rules and actions can be triggered to transform the input data into a form that can be more effectively used, either to standardize the representation (presuming a valid representation) or to correct the values (should known errors be identified).

Standardization builds on parsing by combining a library of data domains to split data values into multiple components and rearrange the components into a normalized format that is recognizable by both human and automated consumers. Standardization can also change full words to abbreviations, or abbreviations to full words, transform nicknames into a standard name form, translate across languages (e.g., Spanish to English), as well as correct common misspellings.

Identity Resolution
It is not unusual that data sets from different systems will have different ways to refer to the same concepts such as product or customer. The desire to search for customer information in different systems and exactly match against existing data sets with a high level of confidence might convince someone that a record might not already exist for a customer when in fact it really does. This is a fundamental challenge: being able to compare identifying data within a pair of records to assess similarity between the data. This helps to suggest that the records have data about the same individual or organization, or to determine that the two records do not truly refer to the same real-world entity.

This is addressed using a technique called identity resolution, in which the similarity of any pair of records is scored using weighted comparisons of key descriptive attributes. If the score is above a specific threshold, the two records are deemed to be a match, and are presented to the end client as most likely to represent the same customer. Identity resolution helps link similar records as well as differentiate record pairs truly represent distinct entities.

Merging and Consolidation of Duplicate Records
Identity resolution provides the foundation of a more sophisticated capability: duplicate record analysis and elimination. Similar records in the same data set are probably duplicates, and may be subjected to cleansing and/or elimination. Similar records in different data sets can be linked together and, in certain scenarios, merged into a single surviving record.

Data Enhancement
Data enhancement is a method of appending extra information to existing records. Data enhancement relies on parsing, standardization, and record linkage from alternate data sets (such as name standardization, demographic data imports, psychographic data imports, and household list append). For example, a frequently-employed enhancement activity is address standardization and cleansing, which relies on parsing, standardization, and the availability of geographic information and associated standards.

Data Cleansing
If it can be determined that there are errors in a record, the data can be subjected to a process of data cleansing, in
which the values are transformed into ones that meet user expectations. Data cleansing employs all of the techniques already described: parsing and standardization, identity resolution, record linkage and enhancement. Data cleansing uses business rules to change incorrect values to correct ones, infer correct data values, correct names or addresses, eliminate extraneous and/or meaningless data, and even contribute to the merge process for duplicate records.

**Steps to Take Action**

Utilities are investing in analytics, both for traditional reasons such as customer experience management, and in preparation for transforming business needs. However, the ability to employ analytics to drive business improvements must rely on good data management practices, especially for data quality assurance. Some initial steps to take involve evaluating organizational capabilities for data quality management, including asking some specific questions:

- Does our organization have a team dedicated to data quality management?
- Have we characterized the dependence of corporate value on information?
- What policies and practices have been institutionalized around data quality assessment, monitoring, and improvement?
- Do we have criteria for assessing compliance with data quality requirements?
- What tools and techniques are in place for assessment?
- Have we assembled the components of a data quality tool kit including parsing and standardization, identity resolution, duplicate elimination, enhancement, and cleansing?

If the answer to any of these questions is not “yes,” the next step is to develop a comprehensive data quality management plan to support your analytics endeavors.
Every connection is a new opportunity™

UNITED STATES
One Global View
Troy, NY 12180
1.800.327.8627
pbsoftware.sales@pb.com

CANADA
26 Wellington Street East
Suite 500
Toronto, ON M5E 1Z2
1.800.268.3282
pbsoftware.canada.sales@pb.com

EUROPE/UNITED KINGDOM
Minton Place
Victoria Street
Windsor, Berkshire SL4 1EG
+44.800.840.0001
pbsoftware.emea@pb.com

ASIA PACIFIC/AUSTRALIA
Level 7, 1 Elizabeth Plaza
North Sydney NSW 2060
+61.2.9437.6255
pbsoftware.australia@pb.com
pbsoftware.singapore@pb.com