Location Intelligence
Location and Business Data

Real-time customer information data quality and location based service determination implementation best practices.

Real-time customer information processing challenges and opportunities

Many organizations need to create and maintain a consistent, single view of customers across the enterprise with accurate and complete customer information. For example, to effectively acquire, serve and grow customer relationships, home insurance providers need to assess the risk for a property and telecommunication providers need to determine if they are adequately meeting customer service level agreements, within specific territories. Poor quality customer information can be costly and will affect sales, billing and cash flow, service delivery and customer relationships.

Jeremy Peters
GISP, CIMP – Master Data Management & Data Quality Solutions Architect / Principal Consultant
Pitney Bowes Customer Information Management and Location Intelligence

Jeremy Peters (GISP, CIMP) is a Principal consultant with Pitney Bowes digital commerce Professional Services team. He has over 20 years of experience in Professional Services as developer, analyst and architect, specializing in the application of Location Intelligence and Data quality technologies for customers across various industries.

A customer data management service is needed to accurately identify and locate customers with geographic precision in relation to the business network. Accurate data and real-time analysis can help ensure that customers receive information regarding the availability of different service offerings. In some cases, the distance from a network asset or the service area in which a customer is located (e.g. electric utility substation, telecommunications wire center, insurance flood zone) must be determined with high geographic precision to ascertain what kind of service the customer is capable of receiving. Organizations need the ability to understand spatial relationships with confidence and make that information available to sales agents, service reps and customers in real time.

The integration of enterprise data quality, address management and location intelligence service-oriented technologies are enabling customer information management (CIM). For example, integration provides the means to correct, standardize, and validate customer information as well as to connect those customers geographically to the business network in a real-time environment using web services. Organizations are also able to determine the serviceability of a customer's location to help sell Services, assign pertinent information to a customer's location required to enable service activation and provide services. These technologies are often implemented using customized rules that are published as Web-services using modern service-oriented architectures (SOA). As such, web services are often easily integrated into new or existing customer management and billing systems.

This article describes best practices for the design and implementation (CIM) of real-time web services. They are presented in the context of an example use case for real-time CIM that is implemented using Pitney Bowes Spectrum Technology Platform software. This use case demonstrates real time telecommunications company customer information and address correction, standardization, validation and geocoding, as well as location based service risk information assignment and exception handling in order to help ensure data quality, data governance and accurate service prequalification and delivery. This article also describes how an advanced mapping application can be used to provide advanced mapping and spatial analysis capabilities in conjunction with a web based portal to help review and analyze exception records where the location precision requires manual review because it is within the error of possibly assigning inaccurate service information.

Use case: real-time customer information correction, standardization, validation, geocoding and location based service information assignment

The Spectrum™ Technology Platform is service-oriented architecture (SOA) platform with a server that supports modules for a broad set of capabilities for data quality, master data management, analytics, Location Intelligence and data integration. These modules provide different functions, such as address validation, geocoding, matching/deduplication, advanced parsing, routing, spatial analysis and mapping among others. The Spectrum Enterprise Designer client tool gives developers drag-and-drop capabilities to construct business-process data flows in the form of batch jobs or web services on a workflow designer.

A web service was developed using the Spectrum platform, to standardize validate and enhance customer information in real time to help ensure customer information data quality, data governance and accurate service delivery. In this case, wireline, wireless and cable providers, who continue to roll-out and maintain telephony services, are required to provide E911 services. E911/Public Safety Service Planning applications assist with network planning and provisioning to help meet Federal Communications Commission (FCC) 911-regulatory compliance. Telecommunication carriers must accurately and efficiently route E911 calls to the appropriate Public Safety Answering Point (PSAP) in order that public safety agencies can dispatch emergency personnel based on the caller location.

These applications provide carriers the information to quickly determine the 911 call centers that responds to emergency calls at a particular location. These types of applications are invaluable for carriers especially as nomadic calling becomes common place. By identifying a customer's location, carriers can accurately provision the user's device to reflect the responding PSAP. (See figure 1.) Through these solutions, carriers can quickly and efficiently determine the serving PSAP for a given location — an address, cell tower, market, coverage area, etc.
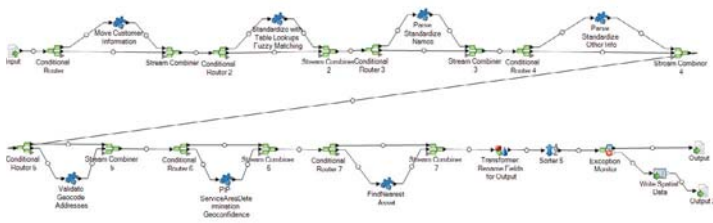
Figure 1. The customer information data quality data flow corrects, standardizes, validates and enhances customer information in real time to help ensure customer information data quality, data governance and accurate service delivery.
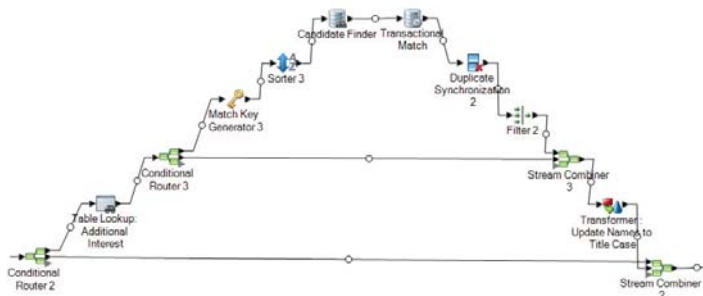
Figure 2. The customer information data quality matching sub flow standardizes company names against a third party and/or internal business list using table lookups and fuzzy matching.

This web service provides the option do any or all of the following:

• Identify and move customer information entered in incorrect fields to the correct fields. For example, the web service identifies parses and moves address information incorrectly entered in personal name fields into appropriate address component fields (e.g. addressline1, addressline2, city, province, postal code).

• Parse and standardize company names against a standardized list of company names. Business name suffixes such as "Incorporated", "Inc.", and Inc" are standardized to one form and parsed from the business name before standardizing company names against a third party and/or internal business list using table lookups and fuzzy matching. (See Figure 2.) The fuzzy matching uses customized matching rules with Phonetic and Edit Distance-based algorithms among others (e.g. Sylable Alignment, Acronym, exact match, Character Frequency and Edit Distance) to match suspect company

names against a candidate group of company names selected to have likely matches from database using a Match Key. Similar matching methodologies are often used to standardize and/or identify duplicate customer information for many elements of customer information, such as personal names, phone numbers, email addresses, dates (e.g. DOB) and identification numbers (e.g. SSN) when sources of candidate information exists to compare against.

• Parse, format and standardize personal names according to business specific rules for handling naming conventions such as conjoined names and natural or reverse order personal names. Personal names are evaluated for gender, order, and punctuation using pattern recognition and name database files. Name data is parsed and concatenated in a new field according to standard formats defined in the business's data quality rules.

• Parse and standardize customer information, such as telephone numbers and emails in order to put it in the same consistent format. This also helps identify if there is a problem with the content if it can't be standardized into the required format.

• Parse, standardize and validate addresses according to postal authority standards. The standard validated address and validation result indicators are returned in order to evaluate the success and quality of the postal address validation. Individual parsed address components, postal information, street name aliases and standardized/normalized input address elements are also optionally returned.

• Geocode customer addresses to the best level of location precision available for each address. In most cases this level of precision is the exact property location. The standardize address, latitude/longitude coordinates and geocoding result indicators are returned by the Geocode Address data flow stage in order to evaluate the success and quality of the address standardization and geocode.

Determine the required service information (Public Safety Answering Point Agency, Flood Zone etc.), found at each customer location using a spatial query, whether it is based on a geocode for an address or the latitude and longitude coordinates acquired by location aware devices like mobile phones.

A boundary/confidence surface is automatically generated around the customer location that represents the possible error in the location coordinate determination depending on how the location coordinates were acquired (e.g. geocoding precision level, GPS, Cell ID, WIFI). (See Figure 3). In this way, the percentage of the 'confidence surface' that overlaps each boundary area can be calculated and quantitative confidence levels can be assigned for the submitted location (refer to Figures 4 & 5). Hence, it can be determined whether the location precision requires manual review because it is within the error of possibly assigning inaccurate service information as a result of the confidence surface overlapping multiple service boundary areas. In order for the call to be assigned to the correct Public Safety Answering Point (PSAP) agency it is important to know If the error built in to the determination of the callers location could place that caller within the boundaries of more than one PSAP agency and what is the probability that the location is within each of the possible PSAP boundaries.

Spectrum technology's "built-in" Find-the-Nearest capability is used when the closest X number of assets (e.g. fiber access points, fire station etc.) within a maximum search distance is used to determine serviceability.
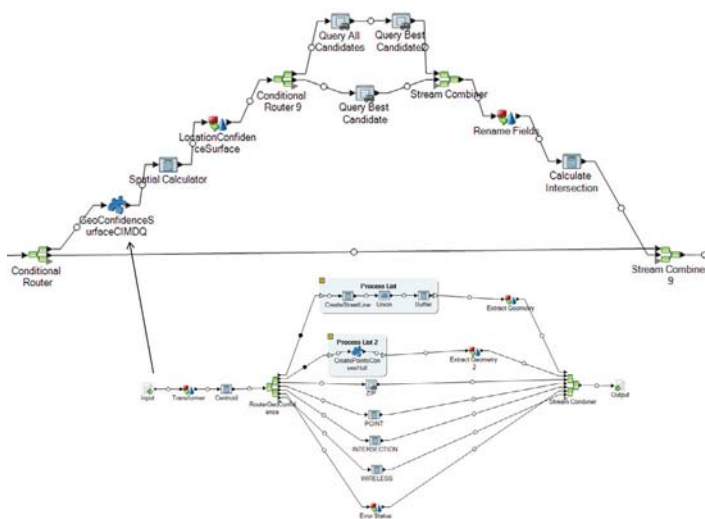
Figure 3. The customer information data quality geo-confidence sub flow generates location "confidence surfaces" that are used to calculate and assign quantitative confidence levels for the submitted location so that it can be determined if the location precision requires manual review because it is within the error of possibly assigning inaccurate service information.

The customer information exception handling capabilities include:

• Route exception records that the web service could not automatically process with a high level of confidence to a web based manual review portal and mapping application for review and correction by a business steward. The web service is configured with a set of conditions for personal name, email, phone, and company name standardization, as well as address validation, geocoding and service determination in order route suspect records to the exception repository and portal for manual review. For example the Service determination condition will route records to the portal if the service area overlaps less than 95% of the confidence surface area and therefore needs to be reviewed to ensure that the correct service information is assigned to the customer. The exception record locations and their confidence surfaces can be viewed in relation to the service area and property boundaries on a map. The exception record locations are color coded on the map according how the location coordinates were obtained (e.g. GPS, Cell ID, WIFI, property precision geocode, street interpolation geocode, ZIP code level geocode) and therefore how the confidence surface was generated. The business Steward can then manually correct any of the information, and optionally automatically submit the record for reprocessing to an exception reprocessing service, or approve the existing exception record as is.

Data quality and Location Analysis best practices

Some of the key customer data qualities best practices employed in this implementation are as follows:

Standardization
Standardize data of the same type in order to put it in the same consistent format. Some types of data that may be standardized include telephone numbers, dates, names, addresses, and identification numbers.

Parse and standardize data before performing matching or deduplication activities because standardized data will be more accurately matched than data that is inconsistently formatted. Inconsistent use of terminology can be a data quality issue that causes difficulty in parsing, lookups, and fuzzy matching.

For example, in this implementation business name suffixes such as "Incorporated", "Inc.", and Inc" are standardized to one form and parsed from the business name before standardizing company names to one form using table lookups and fuzzy matching. The advanced fuzzy matching process maximizes matches and reduces false positives based on matching individual elements such as company name and suffix.

Matching

It is best to start conservative and loosen the matching rules over time in order to avoid false positives records that are incorrectly matched. A wide variety of effective identifiers should be used in matching, such as different types of fuzzy matching algorithms, along with computing weights for each identifier based on its estimated ability to correctly identify a match or a non-match. These weights are then used to calculate the probability that two given records refer to the same entity. For example, compared company names with probabilities above a certain threshold are considered to be matches, while pairs with probabilities in another lower threshold are considered to be possible matches which should be sent for automatic review by the Data Steward.

Define logical candidate groups of records that will have possible matches within them using a match key to match each suspect record against to obtain acceptable matching performance. Match keys facilitate the matching process by allowing you to group records by match key and then only comparing records within these groups instead of potentially comparing millions of records most of which may not be a match. Fields requiring an exact match should be included in the match key and use an appropriate key, such as phonetic algorithm, for the match rule that requires a phonetic match. Matching rules should come from or at least be approved by subject matter expert.

Address validation and geocoding

Accurate address validation matching results are more often achieved when the address is parsed into component fields (AddressLine1, AddressLine2, City, State/Province, Postal Code, Country) and standardized before being submitted to a postal standards based address validation service. Auto-complete address web services can help a user select a suggested address, as the address is typed into an online form, and submit a quality parsed address for

more accurate postal based address validation and/or geocoding. Moreover, a postal based address validation web service can be used to parse and standardize addresses to improve matching and accurate geographic coordinate assignment by a geocoding service.

A greater percentage of addresses can often be accurately validated and geocoded using multi-pass address validation and geocoding where you validate and geocode once with stricter match settings to reduce false positives, but still match as many records as possible. Then validate and/or geocode a second time with less restrictive match settings if you did not get a match the first time. False positives can occur when the geocoder:

• finds a street that sounds like the input street

• finds the same street in another city
  (if postal code match is not required)

• finds the street but with a different house number
  (if house number is not required)

For example, if your address validation processing could not validate the input address, the same web service may be able to with different match settings. In addition, a different address validation or geocoding web service may be able to make a match in a second pass and if not return a list of potential candidate addresses to choose from.

Similarly, with some geocoding web services you can take advantage of cascading match logic combined with multiple conflated/combined datasets to geocode at the best level of precision (Point level, Street level, ZIP+4, ZIP). These advanced geocoding web services use address point interpolation to improve upon regular street segment interpolation by using point data. When an address point dictionary is present an address can be geocoded at the correct property location in most cases and if not at the approximate location on a street segment.

Sorting the address records can reduce processing time when processing micro batches of addresses through a web service. Address records should be by sorted by Country (when processing multiple countries), Postal Code, State/Province and City.

LIM geo-confidence

The probability that a location (latitude and longitude) is within a given area should be determined if the geographic precision of the location has a level of error that is significant in relation to the business need to determine the accurate area information. Regardless of the quality of the address match or location determined by location aware devices like a mobile phone, certain degrees of variance (e.g. address, Cell ID, WIFI, GPS) are incorporated into estimating the longitude and latitude of a location. A dynamic boundary should be created around the location that represents the confidence or error surface for that location so that the percentage of the 'confidence surface' within each service boundary can be calculated and quantitative confidence level assignments can be assigned for the submitted location.

Parcel precision geocoding and Geo-confidence determination used with accurate serviceability data provide the most accurate, location-based service information assignments available. It is important that any location coordinates (Geocoding, GPS etc.) and maps used in this analysis should be analyzed in the same geographic coordinate system (e.g. Longitude/Latitude WGS 84 or NAD 83, Popular Visualization CRS) for accuracy.

Data governance

Exception handling should be built into the web service if any of the processing fails to meet standards. The exception handling should route records that could not successfully process (e.g. address validation or geocoding failures or low confidence matches) to a manual review web portal for review and correction by the data steward. Tools should be available through the web portal to help the data steward fix the issues and then automatically submit the fixed records back to the original web service for re-processing. This exception handling should be part a robust data governance program where controls and processes are established that ensures that important data assets are formally managed throughout the enterprise.

Results and benefits

For many organizations, sub-standard data quality and analysis leads to customers abandoning the process, regulatory fines, poor customer experience and higher churn rates. With the right capabilities, however, organizations can lock-in customer relationships, tackle the high cost of fraud and cross-sell far more effectively. The integration of Enterprise Data Quality, Address Management and Location Intelligence service-oriented technologies can enable better business analysis and decision making by enhancing real-time operational systems and work flows.
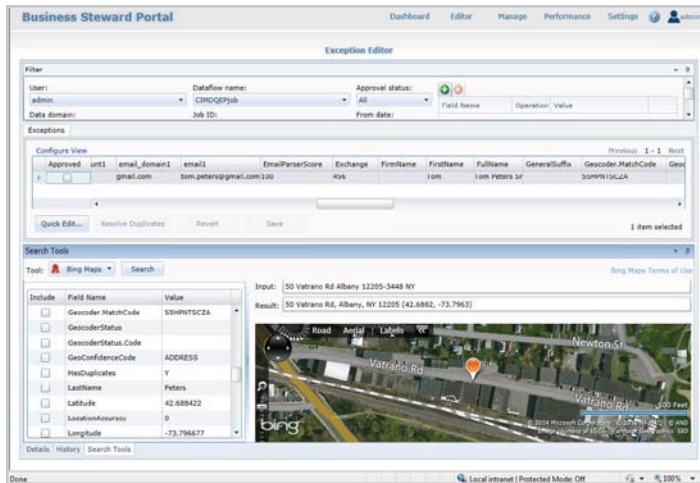

Figure 4. The spectrum business steward module web based portal provides a set of tools to manually review, optionally correct, resubmit for processing and/or approve records that could not be processed with a high level of confidence.
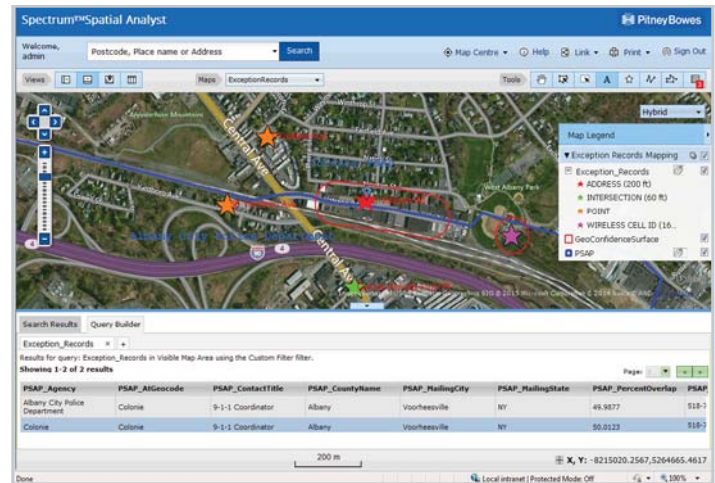

Figure 5. Spectrum spatial analyst application in this case compliments the business steward module by enabling the exception record locations and their confidence surfaces to be viewed in relation to the service area and property boundaries on a map.

pitney bowes

SpatialPoint
www.spatialpoint.com
info@spatialpoint.com • 866-846-5900

For more information,
visit us online: pitneybowes.com

RECYCLE
PLEASE
recycleplease.org

15DCS00971_US