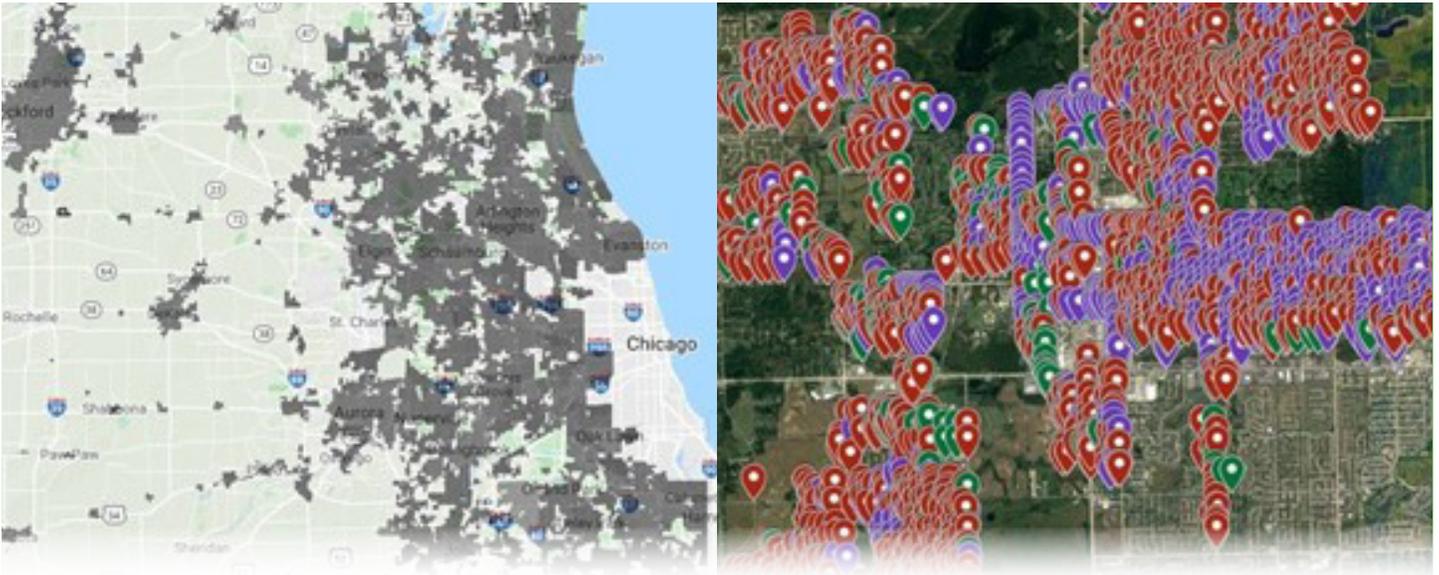


FIELD DATA COLLECTION & MANAGEMENT BEST PRACTICES

By: Shannon Hackett and Andrew Wahrer



INTRODUCTION

Accurate field asset data is critical for utilities to properly maintain and collect revenue. Members of Primera's Consulting team have compiled best practices and lessons learned from field data collection of utility-owned streetlight assets throughout 400 political jurisdictions over three years. Through our experience we've compiled the ten best practices for ensuring data quality and control and will discuss the data collection and maintenance procedure to follow for success. These lessons learned will shed light on the importance of identifying issues early and often to avoid future data errors and laborious manual effort.

Field data collection and management is the act of capturing physical asset information in the field, and a two-way flow of the data to the office and back to the field. There are three major phases of field data collection and management: pre-processing, field data collection, and post-processing. Traditionally,

field data collection has often been treated as the most important phase, but through our experience and analysis we will explain why the emphasis should be put on the pre and post processing of the data. The best practices outlined below will ensure accurate and valuable data is collected for current and ongoing use.

PLAN AND APPROACH

Define Scope

What are you working with and what are you trying to achieve? A clearly defined scope sets expectations, aligns departments, and provides guidelines to measure against. This is especially important if you are working with data that has 100's of different data attributes. It needs to be clearly determined ahead of time what needs to be captured in the field to determine the scope of work and level of effort.

The team's scope was to capture over 60 attributes for each publicly-owned streetlight throughout 400 political jurisdictions using two

existing systems of record. After analyzing the existing data, the team developed an approach to collect, verify, reconcile, and update streetlight information.

Analyze Existing Systems and Data

Know what you are working with: Conduct a data assessment on the existing information to identify gaps, weaknesses, and strengths. A gap assessment can give clarity on where the effort is needed and will guide any data cleanup necessary.

In the example below, the utility had two data sets (pole data and lighting data) they were using to track streetlights. These data sets were not in sync and each held different information.

The data assessment results were as follows:

STREETLIGHT DATA			
ROW		POLE DATA	LIGHTING DATA
1	Streetlight Owner	25%	100%
2	Account number	0%	100%
3	Pole Number	100%	10%
4	Latitude / Longitude	100%	20%
5	Address	0%	90%
6	Lamp style (acorn, cobra)	0%	90%
7	Light type (LED, HPS)	25%	90%
8	Wattage / Lumens (50 watt)	25%	100%
9	City Name	100%	100%

What we Learned

Pole data:

- Pros: Accurate pole location information (row 4), unique identifying number to each pole (row 3).
- Cons: Poor lighting information (rows 6-8).
- Outcome: This data can tell us where the poles are, but can only provide with 25% accuracy which poles have lights attached.

Lighting data:

- Pros: Detailed list of light locations (row 5) with the wattage, bulb and fixture type (rows 6-8).
- Cons: Misleading location information – address could be an actual address (101 Main Street) or a vague description of where the light was (50ft S of Main St alley). Less than 20% of locations had a lat/long which sometimes conflicted with the address (row 4).
- Outcome: This data can tell us the quantity and type of lights that belong to a specific customer but cannot reliably tell us where the light is located.

The results revealed that both data sets had useful information that would need to be linked to provide a clearer picture of what lights were out there and on which poles they were on. The results also revealed that there were still gaps and inconsistencies in the data that would need to be cleaned and validated in the field.

Data Cleanup and Retention of Existing Data

The team determined that it was important to keep as much of the light data as possible. But to be useful, the lighting data set required cleaning. Cleaning in this instance meant that each address in the lighting data set needed to have a corresponding lat/long to map the information. This entailed deciphering address descriptions into an actual address and then assigning a lat/long to that address. Depending on the address description this could take 1-5 minutes per light. The project schedule did not allow for this time-consuming data cleanup and it was bypassed for the first six months of the field data collection project. The lighting data set was not used to its full potential, causing the field crews to miss 10% of the existing streetlights and lose legacy data.

Once the lighting dataset was cleaned and had useful location information, those locations could be validated in the field. The field validation confirmed whether the lights existed in the field, updated attributes for lights that did exist and captured new lights that were missed in the lighting dataset.

Preparing for field data collection

Validating and collecting a large volume of asset data from the field using paper documents to capture information is inefficient and prone to error. There are a wide range of field data collection tools available on the market – these tools usually have an app and web-based version with flexible capabilities. Some capabilities to look for include the ability

to take and store photos, scan QR codes, provide GPS coordinates, and capture attributes in various formats, to name a few.

After choosing a tool and identifying attributes to be captured, you will need to load all existing data into the tool. It's helpful to break the data into approachable sections, such as community or region, but still have the ability to view the data as one piece to ensure locations on the borders of a community or region aren't being captured twice. See 'reconcile with historic data' for additional details.

Post Field Data Collection

Make sure to review and validate before integrating with existing systems. Identify how the data is going to be used and work backward to identify metrics that can be put in place to ensure quality in those areas. For example, if a customer is going to be billed for the wattage of the lights then it would be advantageous to develop processes that check the accuracy of the wattage captured. The next section will highlight the controls that can be put in place to limit human error and ensure the quality of the data.

The benefits of field data to all downstream applications and end-use is dependent on the quality of the data going in. International Organization for Standardization (ISO) quality management standards require Quality The benefits of field data to all downstream applications and end-use is dependent on the quality of the data going in. International

DATA QUALITY ASSURANCE AND QUALITY CONTROL BEST PRACTICES

Organization for Standardization (ISO) quality management standards require Quality Assurance - measures taken to prevent errors from occurring, and Quality Control - steps to detect and correct errors to ensure requirements are met.

Preventing errors during the initial field data collection process is more efficient than attempting to discover and correct errors in the data during post-processing. Additionally, implementing a process to find all errors after data collection is likely not feasible due to schedule and budget constraints. In addition, after the errors are discovered, the revised data also requires a quality control review.

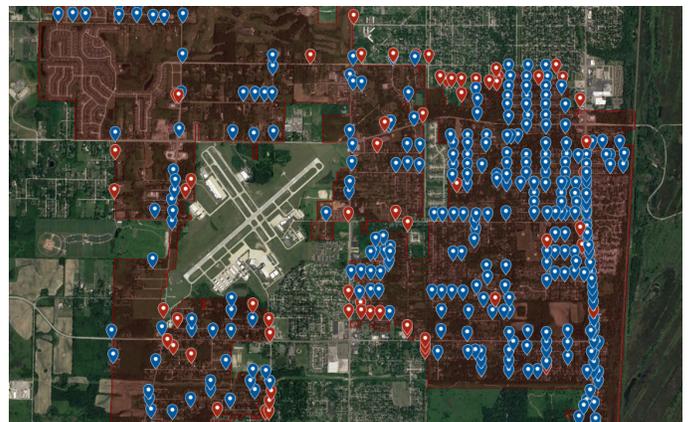
Although error prevention is ideal, errors will inevitably still occur, and a quality control process will be needed. The human error rate on manual data entry alone is 1.2% (1). Considering that it is not reasonable to expect to collect perfect data for every data attribute initially, it may pay off to put supplementary measures and training in place to ensure a higher accuracy rate for data points that are of higher importance.

A quality control program should include error definitions, procedures for identifying errors, data correction procedures, reporting requirements on error rates, and adjusting

procedures to continuously improve the data collection and quality control processes. The following ten areas of best practices will help ensure successful prevention, detection, and correction of data errors in field data collection:

01. Predetermine Geographic Area Boundaries

There are multiple challenges encountered when assigning area boundaries to field data. Streetlights do not have traditional street addresses; their geographical location is best recorded with latitude and longitude coordinates to a high degree of accuracy. Additionally, political geographic boundaries change over time and municipal boundaries may overlap with township and/or county boundaries. To make matters more complicated, some municipal boundaries may be divided parallel down the middle of a street. Implementing preventative measures such as starting the data collection process with a systematic approach that doesn't treat each area boundary independently, and starting with the most up-to-date boundaries from reliable sources such as directly from the municipality will reduce post-data collection work significantly. Using GIS analysis software to assign or correct geographic boundary mapping removes potential for human error when conducting post-data collection analysis.



02. Stakeholder Agreement on Data Format

Over 60 attributes were collected and identified for each streetlight. Agreement on fields and specific data formats should be reached before field data collection begins. Reverse engineering data needs from the ultimate end usage will provide a starting point for data requirements and format. Each system's requirements will need to be analyzed. There should be a primary key data field amongst the systems and there should be a unique identifier for each surveyed asset as well. For instance, throughout the various legacy systems, the primary key for identifying streetlights was a pole number. However, this resulted in various issues; some poles had more than one streetlight which resulted in duplicate IDs assigned to multiple lights, and the pole numbers did not match up in all systems. A separate unique ID was created exclusively for field survey collection data to keep track of the assets while the legacy data and systems were cleaned up.

03. Make Data Collection Simple

Field data should be collected with an application that limits human error and misidentification as much as possible. Simple measures such as avoiding free text entry as much as possible will reduce misspellings and inconsistent data formatting. Dropdown lists should be utilized instead. Inaccurate physical attribute identification is the most difficult to detect during post-data collection quality review. Implementing dependent field controls will help reduce these errors. Dependent field controls filter and limit the values of one field based on the value selected in a previous

control field. For instance, if a streetlight is identified with a LED luminaire, the list of available wattages will be filtered only on those applicable to LED luminaires. Collecting a photo of the asset is paramount considering that data collection errors will still occur even if thorough measures have been taken to prevent them. A photo will allow the quality control team to accurately correct data errors without taking a trip back out to the field.

04. Avoid Duplicate Data

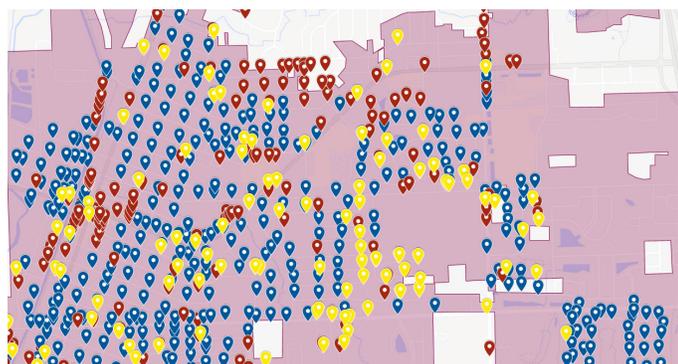
Duplicate locations and missing locations are difficult to identify simply by looking at raw data without GIS analysis. Because streetlights do not have traditional street addresses, duplicate locations may not have been recorded with the exact same address. Latitude and longitude coordinates may vary slightly as well. Using GIS analysis software can easily identify duplicate or expected locations that are missing. Streetlights are typically 50 feet apart, so anything less than that may be flagged for additional review as a potential duplicate. Collecting duplicate and missed location data can be mitigated by planning the field data collection process using a systematic geographic approach that does not silo the geographic scope into separate entities.

05. Reconcile with Historic Data

Reconciling with historic data provides an opportunity for a sanity check against historic records. Even if historic data is known to be inaccurate, variations of significant orders of magnitude are signs that further examination is needed. Utilizing existing data as a starting point for locating assets is important, but also

utilizing existing data to know where not to go is beneficial. For streetlights, the utility only wanted to capture utility-owned locations. However, municipal-owned streetlights often look identical to utility-owned. Utilizing data for municipal-owned locations can cut down on field survey time and data accuracy by identifying locations to avoid.

It may be ideal to start with a fresh slate and replace all existing data with updated survey data at once. However, the utility needed to begin installing new smart LEDs before the entirety of the assets could be surveyed, so new data was integrated with existing data piecemeal. This is where it was critical to identify duplicate data when merging the new data into existing databases and several duplicate checks were developed. This is also where it is beneficial to survey a geographical area in a systematic way that minimizes boundary overlap and thus reduces the chance for duplicate locations near and on the geographic boundaries.



06. Flag for Review in the Field

Flagging a location for additional review helps the quality control team focus their efforts where they are most needed. The best time to flag a location for additional review is during the initial field survey. This allows for more efficient data collection. If a field surveyor cannot resolve an issue in the field, they can flag the location for the quality control team to review later. Predetermining common issues to flag gives congruity to the review process. Common flags include asset ownership identification, geographic boundary review, and asset condition warnings. Additionally, automatic flags for review can be put in place post-data collection. For instance, automatically flagging locations within 100 feet of a boundary for additional review.

07. Third Party Review

Soliciting third party review and feedback is helpful for resolving discrepancies that are not able to be resolved by the quality control team or not able to be resolved in a time-efficient manner. For instance, even though the latest available political boundaries were used to determine which geographic area the streetlight served, in some cases, the streetlight was directly on the boundary line. In this case, the municipal customer was contacted for their feedback. Municipalities are intimately familiar with their jurisdictions and were able to quickly determine the correct municipality that the streetlight serves. Another piece of information the customers facilitated with in the review was asset ownership. Some utility-owned streetlights that are rented to municipal customers look nearly identical

to municipal-owned streetlights and are intermixed throughout the community. And because many of these streetlights were installed decades ago, historic information is not always accurate.

08. Continuous Improvement

Once the initial quality control process is developed, it should not be set in stone. It should be adapted to project changes or as new information is discovered. It should be continuously improved throughout the life of the project. One way to ensure this happens is to solicit and utilize feedback from everyone involved with the project: end users, installers, customers, etc. Give feedback to the field data collectors on a regular basis and train the workforce in changing work practices. Investigating the root cause of data errors will help prevent repeat errors. For instance, using a method to save time on the front end by defaulting data collection steps in the data collection application may cause data errors that require more work during the quality control process that negate any time saved on front end. This was observed during streetlight data collection; the pole material field automatically defaulted to the last surveyed location value. This saved time during data collection by not having to manually populate the field when similar pole types were clustered together. Consequently, this resulted in field data collectors not being accustomed to modifying the pole material field in the application, and when a new pole type was encountered, the field was ignored and not captured accurately.

09. Set, Define, and Measure Acceptable Error Rates

Set and define specific acceptable error rates before data collection begins. Do not use ambiguous targets – i.e., do not decide that a blanket 5% error rate is the target. Some data points may have more tolerance than others. For instance, location accuracy must be within 10 feet. A 5% error rate may be acceptable for some data points, but if 5% of 100,000 assets are not captured, loss of confidence in the data may occur. It is crucial to continuously measure and monitor error rates throughout the life of the project to identify areas for improvement. It is especially beneficial to measure often and critically in the beginning stages to avoid a snowball effect of issues. Scrutinize all details early on, even ones that may currently not seem significant. Later, those details may be valuable, and if there are errors even in seemingly insignificant fields, confidence is lost in all the data.

10. Adjust as Needed

Adjust and implement new quality control steps as new information is learned. If collecting data on assets that have outdated or no existing data stored, it will be impossible to predict all the obstacles ahead of field data collection. Continuously improving and adapting the quality control process will ensure that quality of future data collection improves.

REFERENCES

1. Smyth ET, McIlvenny G, Barr JG, Dickson LM, Thompson IM. Automated entry of hospital infection surveillance data. *Infect Control Hosp Epidemiol.* 1997;18(7):486-491. doi:10.1086/647653. From <https://pubmed.ncbi.nlm.nih.gov/9247831/>

ABOUT THE AUTHORS

Shannon Hackett

Project Manager Consultant

Shannon Hackett is a project manager at Primera Engineers. Her expertise is the result of five years of experience with municipal engagement, smart technology, and project tracking. Her experience includes extensive project tracking, database management, and data analysis; material procurement and land acquisition for a 60 mile transmission line; and permitting and planning for distribution construction. As a member of Primera's consulting team, Shannon is running streetlight projects and providing her project management skills to drive performance.



Andrew Wahrer

Business Analyst

Andrew Wahrer is one of Primera's expert business analysts. His expertise is the result of over 10 years of experience with improving work practices, executive level reporting, benchmarking multi-billion-dollar investment portfolios, and increasing productivity through analytical insights in the utility, engineering, and marketing industries.

