RESEARCH FOUNDATION

RESEARCH FOR THE NFPA MISSION

TECHNICAL NOTES

Residential Electrical Fire Problem: The Data Landscape

REPORT BY:

Victoria Hutchison
Fire Protection Research Foundation

October 2018

1 Batterymarch Park, Quincy, MA 02169-7417, USA
Email: foundation@nfpa.org | Web: nfpa.org/foundation

-----This page in intentionally left blank -----

FORWARD

The causes contributing to residential fires due to electrical failures is currently a concern within the electrical community. The National Electrical Code (NEC®) focuses on ways to reduce the risks of fires and shock hazards by incorporating new technology and knowledge with respect to electrical safety. However, data and data analytics is lacking to guide the optimum approaches to minimize residential electrical fires and related hazards.

While traditional data collection approaches have shortcomings that make their ultimate value questionable, future data collection approaches are becoming available. These emerging approaches are generating a new perspective to support optimum data collection through the promotion of new technologies, frameworks and protocols that were not previously available. This technical note summarizes the data landscape for residential electrical systems and related fire incidents, identifying key gaps and challenges in the data and data analytics that is preventing the residential electrical fire problem and the impact of the NEC's regulatory changes regarding AFCI's from being precisely defined.

This has been an FPRF-led project with administrative oversight provided by the Fire Protection Research Foundation in accordance with its advisory services policies. The Research Foundation appreciates the guidance provided by the Project Technical Panelists, the funding provided by the National Fire Protection Association (NFPA), and all others that contributed to this effort.

About the Fire Protection Research Foundation

The Fire Protection Research Foundation plans, manages, and communicates research on a broad range of fire safety issues in collaboration with scientists and laboratories around the world. The Foundation is an affiliate of NFPA.

About the National Fire Protection Association (NFPA)

Founded in 1896, NFPA is a global, nonprofit organization devoted to eliminating death, injury, property and economic loss due to fire, electrical and related hazards. The association delivers information and knowledge through more than 300 consensus codes and standards, research, training, education, outreach and advocacy; and by partnering with others who share an interest in furthering the NFPA mission.

All NFPA codes and standards can be viewed online for free.

NFPA's membership totals more than 65,000 individuals around the world.

Keywords: Datasets, Data Collection, Data Analytics, Residential Fires, Electrical Fires, NFPA 70 (NEC), AFCI

Report number: FPRF-2018-11
FPRF Project Manager: Casey Grant

-----This page in intentionally left blank ----

PROJECT TECHNICAL PANEL

Donny Cook, Shelby County, AL Department of Development Services
Mark Earley, NFPA
Joseph Gochal, NFPA Data & Analytics
Mark Hilbert, MR Hilbert Electrical Inspections & Training
Kevin Lippert, Eaton Corporation
Steve Rood, Legrand, North America

PROJECT SPONSOR

NFPA®

-----This page in intentionally left blank -----

Table of Contents

Table of Figures

List of Tables

1. Background

The causes contributing to residential fires due to electrical failures is currently a concern within the electrical community. The National Electrical Code (NEC®) focuses on ways to reduce the risks of fires and shock hazards by incorporating new technology and knowledge with respect to electrical safety. However, data and data analytics is lacking to guide the optimum approaches to minimize residential electrical fires and related hazards.

Today's safety infrastructure incorporates multiple approaches and technologies to achieve safer electrical installations. Understanding the effectiveness of these technologies, arc-fault circuit interrupters (AFCI's) in particular, and their impact on residential electrical fire incidents would be valuable. Data is needed to assist NEC Code Making Panel (CMP) #2 with determining the best methods for protecting branch circuit wiring in dwelling units against electric arcing. But to be useful, specific data elements need to be collected and evaluated. Examples of needed data elements include the age of construction, wiring methods, type of electrical failure, causal and contributing factors, location of the fire, the NEC edition in effect at time of construction, and the branch circuit protection devices installed.

Unfortunately, there are inherent challenges and barriers to the effective collection of the applicable data. Traditional data collection approaches have shortcomings that make their ultimate value questionable (e.g., lack of detail and quality on fire department collected residential fire events). Further, not all existing datasets are openly accessible, is lacking specific important details, or insufficient in quantity.

Despite handicaps with existing data, future data collection approaches are becoming available. Today's world is introducing new opportunities for novel approaches to address these data needs. For example, the evolution of the internet of things (IoT), smart cities, cyber physical systems, and open data sources are supporting new means of collecting data that did not previously exist. This is generating a new perspective to support optimum data collection through the promotion of new technologies, frameworks and protocols that were not previously available.

2. Objective

The overall objective of this report is to summarize the data landscape for residential electrical systems and related fire incidents. This report identifies key gaps and challenges in the data and data analytics that is preventing the residential electrical fire problem and the impact of the NEC's regulatory changes regarding AFCI's from being precisely defined.

This Technical Note addresses residential electrical data through the following:

- Identification of currently available data sets;
- Identification of trends and anticipated and envisioned future data sets;
- Address the positive and negative characteristics of these data sets;
- Summarize the gaps and challenges in the data and data analytics; and
- Provide recommendations for actionable next steps to continue to address the gaps and challenges.

3. Data

According to NFPA's report on *Electrical Fires*, there are an estimated 45,210 residential fires each year in the United States involving an electrical failure or malfunction, resulting in 420 deaths, 1,370 injuries and $1.4 billion in direct property damage annually (Campbell, 2017). Despite the general downward trend of electrical fires since 1980, there was a significant increase in residential electrical fires between the years of 2012 and 2014, as shown in Figure 1 below.

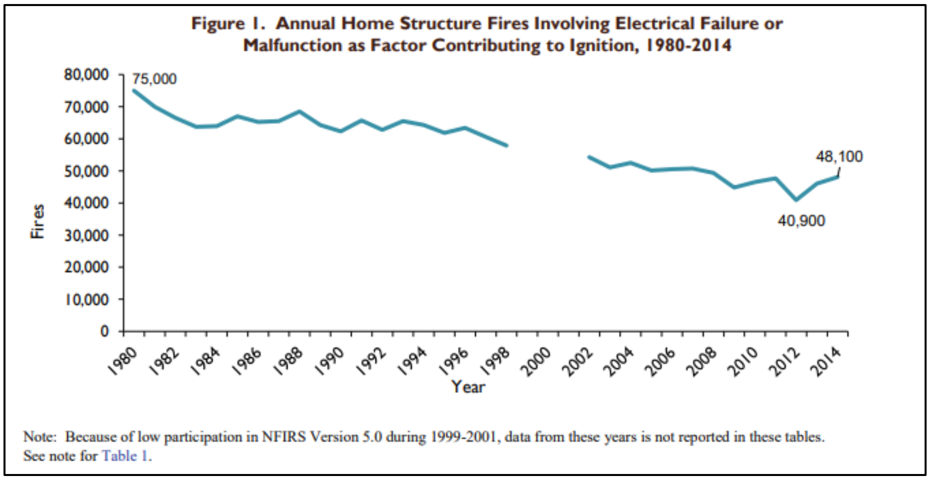

Figure 1 Annual Home Fires Involving Electrical Failure (Image retrieved from Campbell, 2017)

Looking deeper into the types of electrical fires reported, 51% were classified as "unknown electrical" and the remaining percentage were related to some form of arcing in the electrical wiring (e.g. unspecified short circuit arc, short circuit arc from defective or worn insulation, arc or spark from operating equipment, arc from faulty contact or broken conductor, short circuit arc from mechanical damage, or water-caused short circuit arc) (Campbell, 2017).

As shown in Figure 2 below from NFPA's report on *Electrical Fires,* home fires involving electrical wiring and related equipment have generally been on the rise since 2002, as shown in the figure below. Fires specific to electrical wiring and related equipment account for a total annual average of 22,190 fires (49% of total electrical fires), 230 civilian deaths (55% of total civilian deaths), 630 civilian injuries (46% of total civilian injuries), and $804 million in property damage (57% of total property damage due to electrical fires) (Campbell, 2017).

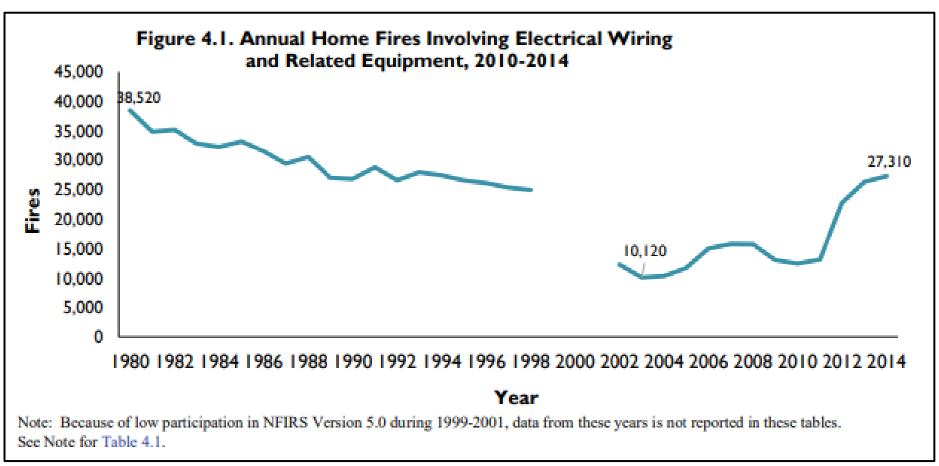

Figure 2 Annual Home Fires Involving Electrical Wiring (Image retrieved from Campbell, 2017)

The increase in residential electrical fires and the alarmingly high percentage of unclassified electrical incidents highlights the need for quality data on residential electrical fires. The similarly high number of incidents relating to electrical arcing also stresses the need for an evaluation of the effectiveness of arc protection devices in residential dwellings. These statistics bring us back to the ubiquitous issue of data.

Currently, there is limited data available on AFCI's impact on residential electrical fires. This is largely due to the fact that fire statistics do not measure the effectiveness of prevention devices. If an AFCI functions properly, it will detect the arc fault, de-energize the circuit, and likely prevent ignition. As a result, there is uncertainty regarding the residential electrical fire problem and the effectiveness of branch circuit protection devices, such as AFCI's.

Therefore, to assess the effectiveness of our current electrical protection methods, data is needed. Once collected, raw data can be refined, analyzed, and given context to develop information and knowledge that can be utilized to inform decisions. With particular reference to residential electrical fires, there is potential to inform analyses of electrical fire incidents and even provide insights into the performance of electrical systems and detection devices through the collection, processing, and delivery of residential electrical data.

3.1 Core Principles for Data Collection and Analytics

According to the world of cyber physical systems, there are three core principles behind the concept of data collection and analytics (Grant, Hamins, Bryner, Jones, & Koepke, 2015):

(1) Collect the data
(2) Process (Analyze) the data
(3) Deliver the data

As shown in Figure 3 below, once the data is collected, processed, and delivered, it can be used to inform decisions regarding electrical safety.

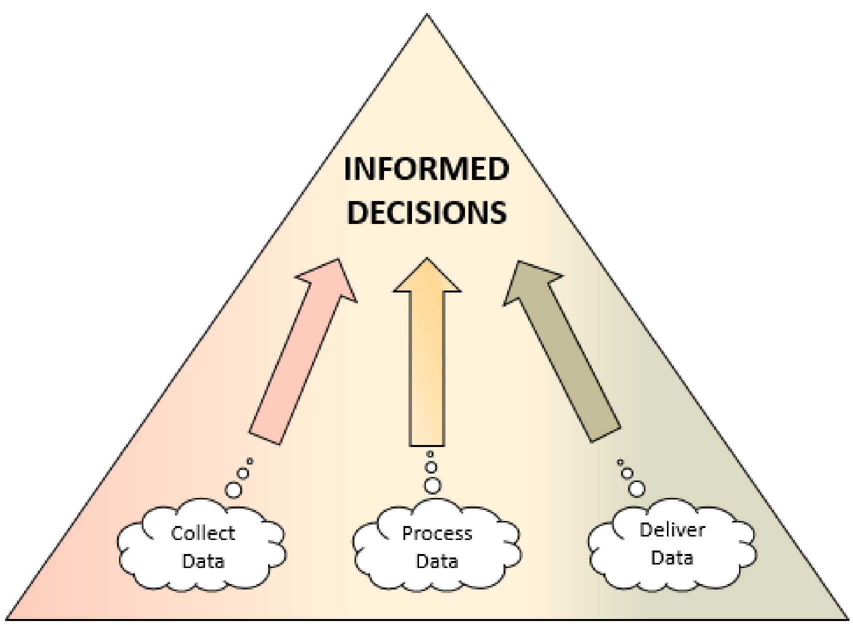

Figure 3 Core Principles of Data Collection and Analytics

3.1.1 Data Collection.

Data collection is the first step to enabling the effectiveness of electrical branch circuit protection methods to be evaluated. Data is collected on the backbone of technology (e.g. software, databases, devices, etc.). The raw data obtained from various sources lay the foundation for data analytics and the development of value-added insights.

To process the data, it is critical that the data collected is of good quality; as poor quality data has minimal value. Data quality is specifically in reference to three characteristics:

(1) *Relevance* – is the data current or out of date?
(2) *Format* – is the data in a format that can be used for data analytics?
(3) *Reliability/Accuracy* – is the data accurate, true, and free of errors?

If good quality data is collected, it is in an adequate state to be processed and analyzed to develop meaningful information and insights.

3.1.2 Data Processing

Data processing or analytics refers to the process in which raw data elements are examined and given context to enable conclusions to be drawn from the information contained within the dataset (Grant, Hamins, Bryner, Jones, & Koepke, 2015). The processing phase includes compiling and integrating data from various sources and giving context and meaning to the data to develop useful information and knowledge to assess the residential electrical fire problem.

3.1.3 Data Delivery

Once the raw, quality data is collected, refined, analyzed, and processed to develop information and knowledge, this knowledge can then be delivered to the applicable party (e.g. person, database, device, etc.) at the time it is needed. After this data evolves into knowledge, it could be applied and utilized to inform decisions regarding the most effective approach to protecting residential electrical branch circuits against electrical arcing.

Thus, on the backbone of technology, we collect data. Processed data yields information and knowledge (Grant, Hamins, Bryner, Jones, & Koepke, 2015). With the right information delivered to the right people when it is needed, comes the ability to inform decisions regarding residential electrical safety.

3.2 Approaches to Evaluate the Protection of Branch Circuits

There are primarily two approaches available to evaluate the effectiveness of arc-fault protection devices for electrical branch circuits: (Approach 1) Collect and analyze retrospective fire incident data on residential electrical fires or (Approach 2) leverage the use of emerging technologies to utilize real-time monitoring of residential electrical systems to collect prospective data.

For a retrospective data collection approach for residential electrical fires, the analysis is limited to the data that is currently available in the various formats. This approach requires multiple inputs, from multiple data sources to be collected for each home evaluated. Once the data is collected, the various data elements of varying quality and format will be consolidated to create a body of data to manually analyze the collected, incident centric data. This approach is graphically illustrated in Figure 4 below and further discussed in Section 3.2.1.

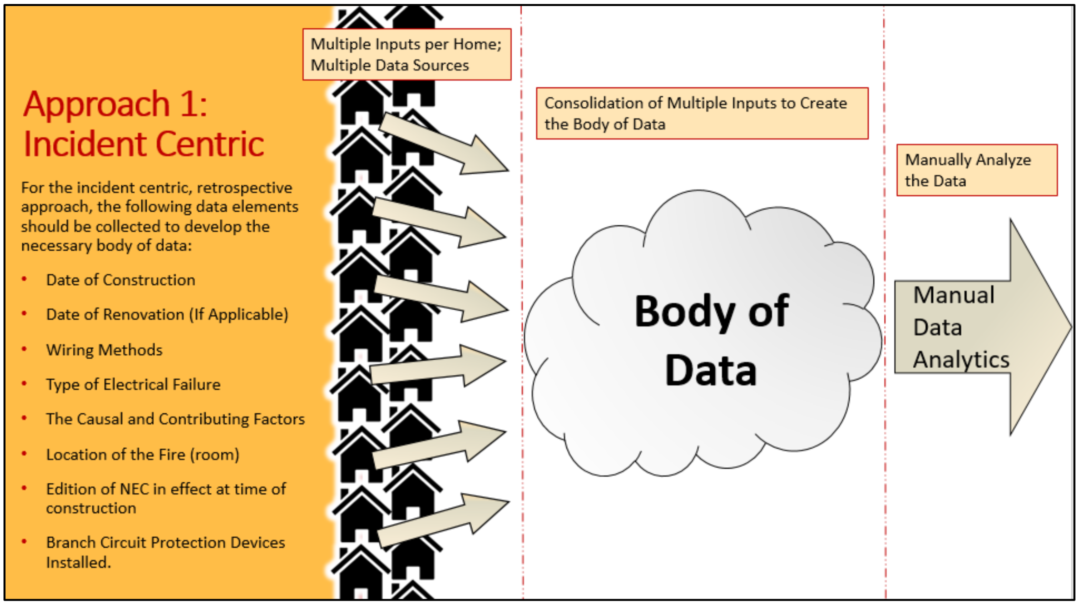

Figure 4 Retrospective Approach of Residential Electrical Fire Incidents

Figure 5 below depicts a prospective, system centric approach. Since a prospective approach to data collection for residential electrical systems focuses on collecting data going forward, there is an ability to guide the outcome of the data collected (e.g. the quality, the format, specific data elements needed, quantity, etc.). This approach takes a holistic view of residential electrical systems. Through the continual monitoring and reporting capabilities of prospective data collection approaches, the effectiveness of electrical protection devices have the potential to be identified. This approach is graphically illustrated in Figure 5 and discussed further in Section 3.2.2.

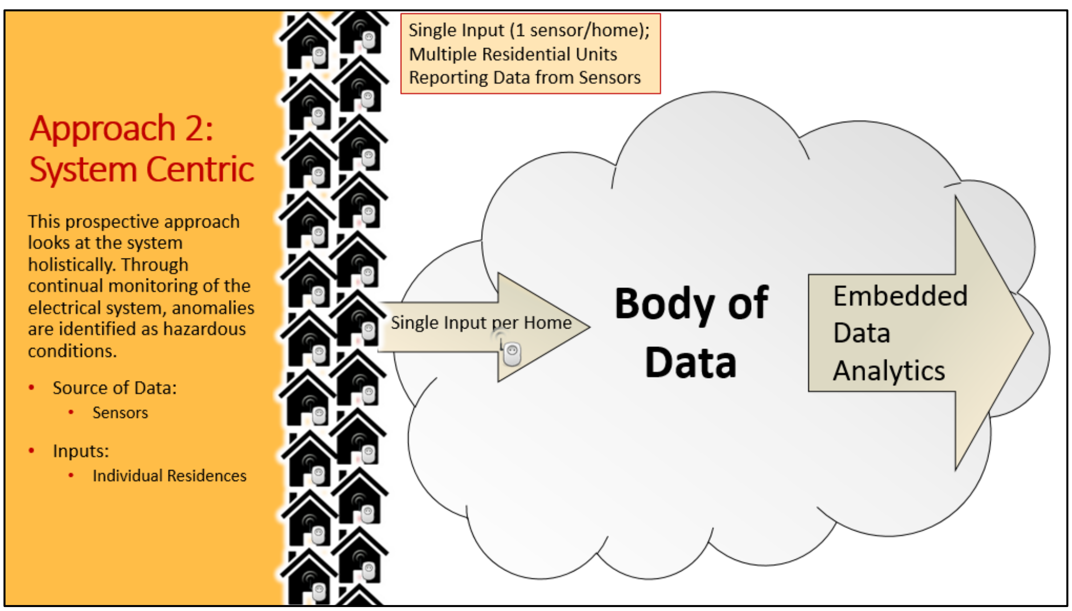

Figure 5 Prospective Approach to Evaluate Protection of Electrical Branch Circuits

The required steps for these approaches are discussed in greater detail in sections 3.2.1 and 3.2.2.

3.2.1 Approach 1: Retrospective analysis of residential electrical fire incidents

To perform a retrospective analysis of past residential electrical fire incidents focused on the effectiveness of AFCI's, the data elements that should be collected, at minimum, for this analysis include the following:

- Date of construction
- Date of renovation (if applicable)
- Wiring methods
- Type of electrical failure
- The causal and contributing factors
- Location of the fire (room)
- Edition of the NEC was in effect for the particular installation
- Type of branch circuit protection devices installed
- Others

The date of construction provides insights into the edition of the NEC was in effect at the time of construction, the age of the electrical system, common electrical installation practices, wiring methods, etc. Since residential occupancies are only required to comply with the edition of the NEC in effect at the time of construction, the date of construction could specify if and where AFCI's were required in the home. This data element will also enable analysis of the residential fire problem regarding where most electrical fires occur (e.g. old versus new construction, with or without AFCI's, etc.).

The date of renovation could reveal the age of the electrical system. For example, renovations can result in upgrading the electrical distribution wiring or adding additional electrical safety devices (e.g. AFCI's). The renovation date can give insight into the age of the electrical system, depending on the extent of renovation. However, caution should be used when equating the renovation date to the age of the electrical system. Just because a home was renovated does not necessarily ensure that the home's electrical system was upgraded to fully comply with current NEC regulations.

The type of wiring is another element which is essential to the evaluation of the cause of the residential electrical fire. For instance, aluminum wiring has primarily been eliminated from use in new installations since it has presented significant fire safety concerns. Although aluminum wiring in itself does not cause a problem, when it is connected to outlets, light switches, or junction boxes, the connection can deteriorate and create a fire hazard (Romano, 2006). When evaluating residential electrical fire incidents, the type of wiring could be a contributing factor to the fire cause.

The type of electrical failure is a critical element to be evaluated. A few common descriptions of types of electrical failure including the following: electrical wiring/equipment problem, electrical arcing, shorted electrical equipment, unknown electrical, electrical distribution failure, electrical malfunction, or other. Whether the electrical malfunction was associated with the electrical infrastructure (e.g. cabling, receptacles, lighting) or plug and cord connected products will typically be indicated (U.S. Fire Administration, 2018).

The cause and contributing factors of residential electrical fires should be collected to identify and understand what equipment failed, what combination of factors contributed to the incident, and the effectiveness of preventative efforts (if appropriate).

One of the most critical data elements to collect is the type of branch circuit protection devices installed in the residence. Home owners may exceed or fall short of the NEC requirements for branch circuit protection devices. Therefore, before the effectiveness of such devices can be evaluated, the type of branch circuit protection devices that are installed and where must first be determined (e.g. AFCI receptacles in specific locations, AFCI circuit breaker, etc.).

The location of the fire within the residential dwelling is also an important data element to collect. The requirement for AFCI installations in homes has expanded over the past several editions of the NEC – expanding the requirement to more locations within the home. This data element will need to be mapped to the type of branch circuit protection devices that were installed in addition to the edition of the NEC in effect at the time of construction to evaluate the effectiveness of the installed protection devices (if any) in the area of origin.

All data elements mentioned herein, in addition to the respective NEC edition in effect, are important to collect for a retrospective analysis of residential electrical fire incidents to provide insights into the effectiveness of AFCI's.

The resources and databases available to collect the data elements mentioned herein are discussed in Section 4 of this report. However, it should be noted that proving the effectiveness of branch circuit protection devices is a challenging task to accomplish using a retrospective approach.

A summary of this retrospective, incident centric approach is provided in the Figure 6 below. Once a fire incident occurs, a fire department responds and manages the incident. After which, a fire incident report is written. When performing an evaluation of the electrical fire problem, a large sample of residential electrical fire incident reports could be collected. These incidents would then need to be supplemented with additional incident centric variables from various sources needed to evaluate the residential fire problem with respect to electrical arcing. (Note that the supplemental inputs indicated in the figure below are intended to provide examples of data sources from which the additional, required data elements could be acquired). All collected data would then need to be aggregated and mapped to each incident to develop the body of data for further analysis. As demonstrated in the figure below, an extensive effort would be required to collect, process, and assess the relevant data residential electrical fire data.

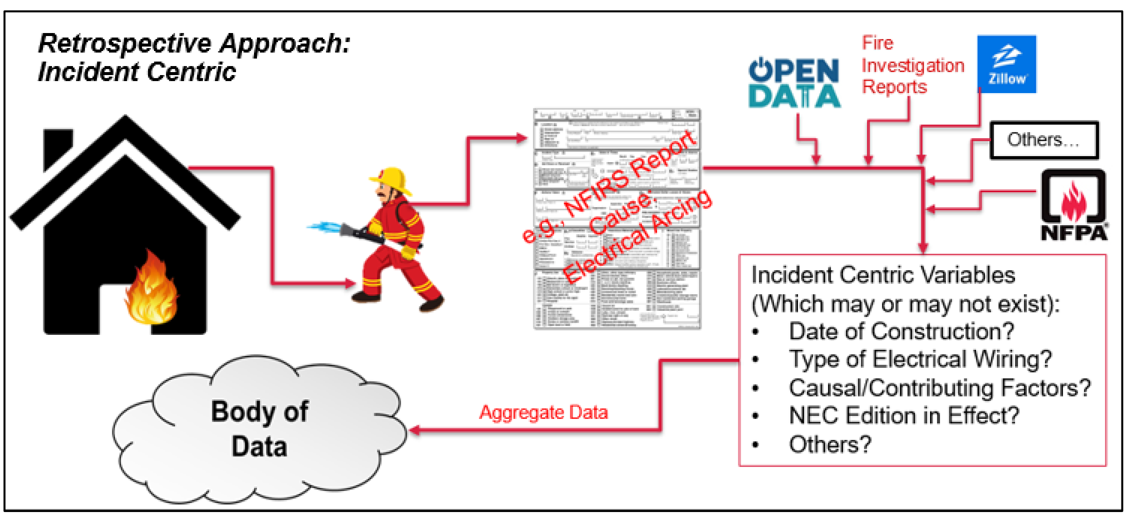

Figure 6 Data collection sequence for a retrospective, incident centric approach

3.2.2 Approach 2: Prospective analysis of residential electrical systems

This approach provides a more holistic view of the entire electrical system. Instead of focusing on post-event analysis, this approach focuses on the real-time status, health, and performance of the electrical system. Leveraging the use of the most current technologies is essential to perform a prospective analysis of residential electrical systems.

The goal of this approach is to prove the effectiveness of branch circuit protection devices, such as AFCI's. When an electrical system is continually monitored, irregularities in the behavior of the electrical system have the potential to be detected, identified, and reported. The presence of hazardous arcs or when an AFCI activates and shuts down the electrical system may be documented in the data output.

Many of the data elements required for Approach 1 are not needed or relevant for this prospective, system centric approach. The prospective analysis is not focused on an evaluation of a post-fire incident, but is instead able to determine whether an AFCI activates when a hazardous condition is present in the circuit by continually monitoring the electrical system's electromagnetic data. Beyond the potential to identify hazardous conditions, the data collected may also be able to provide insights into end-user behavior and activity.

Data-driven approaches have the potential to offer real-time insights into the performance of residential electrical systems and protection devices, while also increasing awareness of hazards within the electrical system. Although significant computing and data storage resources are required for this approach, the effort required to collect, analyze and assess the data is considerably reduced for this approach. This concept is illustrated in Figure 7 below.

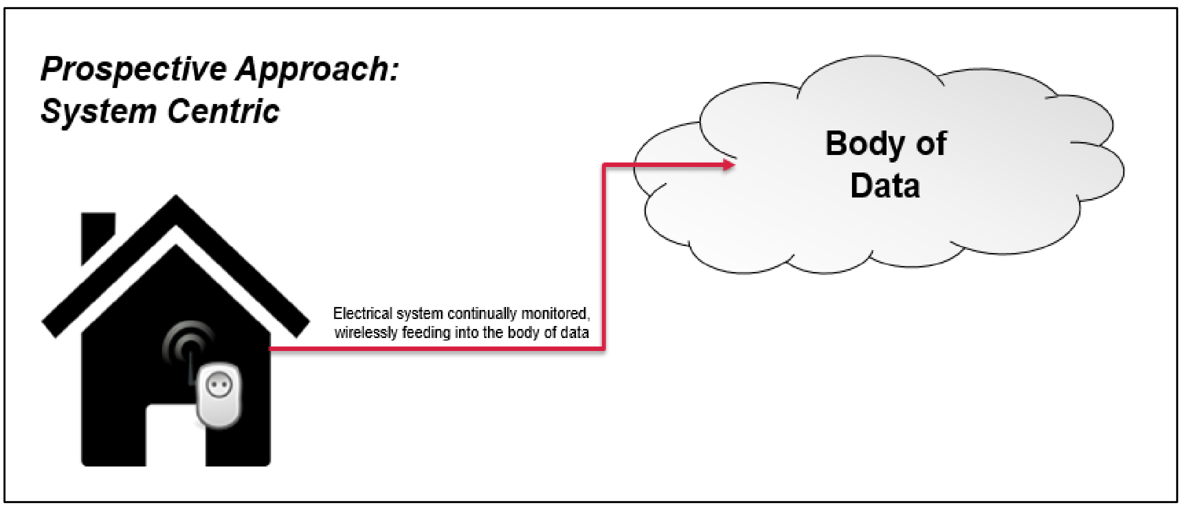

Figure 7 Data collection sequence for a prospective, system centric approach

The potential opportunities for such approaches will be further discussed in Section 5 of this report.

4. Existing Data Landscape

This section provides an overview of the data landscape for residential electrical fires. What data is currently available through various traditional and emerging data sources, the characteristics of each data set, and the data collection and analytic challenges associated with each dataset is discussed herein. The dataset evaluations discussed within this section are specifically focused on the needed data elements and desired characteristics for residential electrical fire data.

4.1 Traditional Datasets

Traditional datasets refer to established databases and practices that are currently implemented as a means of data collection. This section is entirely focused on these datasets with respect to residential electrical fire data (as outlined in Section 3.2.1). It should be noted that other categories of data within the datasets mentioned herein may have different characteristics.

The traditional datasets discussed throughout this section include the National Fire Incident Reporting System (NFIRS), Fire Incident Data Organization (FIDO), and Fire Investigation Reports. The traditional datasets mentioned within this section are not intended to be an all-encompassing list. However, the listed datasets are examples of common platforms for acquiring data on residential electrical fire incidents and thus are the focus of the evaluation of the traditional, existing data landscape.

4.1.1 National Fire Incident Reporting System

The National Fire Incident Reporting System (NFIRS) is a fire incident collection system which was established by the National Fire Data Center of the United States Fire Administration (USFA). The USFA is a division of the Federal Emergency Management Agency (FEMA). NFIRS is currently the largest, annual database of fire incidents in the United States at this time. Although fire departments are not mandated to participate in NFIRS, approximately 24,000 fire departments report approximately 1,200,000 fires annually (U.S. Fire Administration, 2018).

An NFIRS report will include the following modules: the basic module, fire module, structure fire module, civilian fire casualty module, fire service casualty module, EMS module, hazardous materials module, wildland fire module, apparatus or resources module, personnel module, and the arson module. Besides the basic module, only the modules applicable to the specific incident are required to be completed (Thomas & Butry, 2011).

The basic module is required for all incidents that the fire department responds to. This module requests the following information: the incident, incident type, alarm time and response time, actions taken, property losses, casualties, and property use (Thomas & Butry, 2011). The fire module within an NFIRS report requires the cause of the fire, the fuel source, the ignition source, and location of the fire to also be documented which is all relevant to this study.

4.1.1.1 Positive Characteristics

Large sample size. Although it is not mandatory, all fire departments throughout the U.S. are encouraged to fill out NFIRS reports and send them to their respective State's Fire Marshal's Office. Approximately 75% of U.S. Fire Departments fill out and submit NFIRS reports after responding to a fire incident (U.S. Fire Administration, 2018). Therefore, NFIRS reports provide a sample of fire incidents that is likely representative of the entire U.S.

Standardized data collection format. The standardized format provides consistency in the information reported and collected from an NFIRS report. Although the format is not ideal in terms of compiling large quantities of data, the consistency in the type of information and format enables the information to be assessed more easily.

General Incident Information. An NFIRS report provides information on the classification of the incident (e.g. electrical malfunction, failure of electrical distribution equipment, etc.), location, type of occupancy, response information, etc.

4.1.1.2 Negative Characteristics

Lack of granularity. The lack of granular detail in an NFIRS report hinders a detailed analysis of residential fires due to electrical failures. It should be noted that an NFIRS report was not designed to provide granular information, but rather was intended to provide general incident information.

Lack of needed information. Key details such as the construction type, the year the residential structure was built and/or renovated, type of wiring, and branch circuit protection devices installed are not included in an NFIRS report. Although causal factors may be provided, there is often significant uncertainty associated with reported cause of the incident.

Frequent Misclassification. It is common for the cause of electrical fire to be misclassified. This could be due a lack of electrical expertise by the responding fire department. For instance, in NFPA's report *Electrical Fires,* 51% of fires between 2010 and 2014 were identified as having an unclassified electrical failure or malfunction contributing to ignition, which highlights the uncertainty around causal factors of electrical fire incidents (Campbell, 2017).

Inability to use data analytics techniques. The format of NFIRS reports presents challenges for assessing large quantities of residential electrical fire data. The format (standardized form) also hinders the ability to automatically overlay datasets for a comparative analysis. Extracting residential electrical fire data from NFIRS reports, in the current format, can be a time intensive process.

4.1.1.3 Gaps and Challenges in the Data and Data Analytics

Despite its positive characteristics, there are a vast array of challenges presented by NFIRS data. Since NFIRS data is collected on a "per incident" basis, it is difficult to analyze mass sums of data. Although the data is generated on standardized forms, the data is not in a format that can be easily combined with other data sets. While NFIRS data is valuable, it does not provide all the information needed for this analysis.

4.1.2 <u>Fire Incident Data Organization</u>

The Fire Incident Data Organization (FIDO) is an internal database, owned and managed by the National Fire Protection Association, which gathers fire incident data of interest to NFPA (e.g. large loss fires, fires with fire fighter fatalities, fires with specific causes, etc.). Fire incident reports are requested and collected from responding fire departments and stored in the FIDO database. Fire Department participation is voluntary – therefore, the formats of the information collected are not standardized and the documentation can vary depending on the fire department providing the incident details.

4.1.2.1 Positive Characteristics

<u>Additional Information.</u> There is often a narrative and pictures attached in these reports, providing a more detailed description of the incident, cause, description of the building, identification of additional and relevant pieces of information from the responding fire department, etc.

<u>Specify Target Incidents.</u> Since this is an internal database, NFPA has the ability to target specific incidents to get a larger collection of a specific category of incidents, such as residential electrical fires.

<u>Granular Detail.</u> FIDO reports generally contain granular detail in the fire incident reports and narratives which permits a detailed analysis of residential fires due to electrical failures to be completed.

4.1.2.2 Negative Characteristics

<u>Inconsistency.</u> Within this database, the incident reports are requested from various fire departments to contribute to fire incident analysis studies. The response from the fire department is voluntary. Although specific information may be requested, the fire department will often submit their incident report or an NFIRS report. The contents and granularity of the information within the report can vary.

<u>Selective Sampling of Incidents.</u> At this time, residential electrical fires has not been a targeted incident category. So the number of these incidents in the database is somewhat sparse and is currently a relatively small sample of incidents. It should also be noted that the majority of the residential electrical incidents currently in this database are large-loss fire incidents.

<u>Format is not data friendly.</u> The format of FIDO reports presents challenges for analyzing large quantities of data (i.e. FIDO reports). These reports are typically hand written and filled out by the responding fire departments on individual forms. The format may hinder the ability to automatically overlay datasets for a comparative analysis or for use of advanced data analytics techniques.

4.1.2.3 Gaps and Challenges in the Data and Data Analytics

The primary challenge of using FIDO is the lack of incidents specific to residential electrical fires at this time and its format which presents challenges for performing analytics. While the lack of residential electrical incidents is currently a gap, more of this category of incidents could be sought in the future to develop a more robust database of residential electrical fire incidents. The current format makes analysis of this data challenging, however, the use of emerging data analytic techniques such as text mining have the potential to ease this burden in the future.

4.1.3 Fire Investigation Reports

A fire investigation is the analysis of a fire-related incident, after firefighters have extinguished the fire, to determine the origin, cause and contributing factors of the fire. The findings from fire investigations can help prevent future incidents and provide insights into the effectiveness of electrical safety products. The positive and negative characteristics of data from fire investigation reports are discussed below.

4.1.3.1 Positive Characteristics

Large Quantity of Relevant Information. It is common for a fire investigation report to note the construction type, the age of the residential building (if known), type of incident (e.g. electrical distribution fire), fire origin, and the cause (often determined through destructive assessments), path of fire spread, etc. If a destructive assessment is conducted, the wiring methods can often also be determined along with the type of branch circuit protection devices installed. The information (i.e. data elements) provided within a full fire investigation is adequate to evaluate the effectiveness of AFCI's (if installed) on residential electrical fires.

Extensive Detail. Fire investigation reports typically go into extensive detail, including a narrative and pictures of the incident. The quantity and quality of information provided within a fire investigation report performed by an electrical expert would provide the information needed for this analysis.

Good Quality. Fire Investigators go through significant efforts to collect data that is accurate and proven factual through scientific methods.

4.1.3.2 Negative Characteristics

Small Sample Size. Fire investigations are not conducted on every incident. Commonly when there is a large-loss fire in terms of fatalities and property damage, the fire incident will be investigated. The relatively small sample size of investigated residential electrical fires, could be a limiting factor for this analysis.

Time. In order to acquire a significant sample size of fire incidents, fire investigations will likely have to be conducted over a significant period of time. The incidents that can be investigated are dependent on the frequency of residential electrical fires and the availability of the technical expertise when needed – thus, this process of gathering residential electrical fire investigation reports could extend over a significant time period.

Format is not data friendly. Fire investigation reports are typically provided as a narrative. While the granular information is valuable, the format hinders the ability to assess large data sets, consolidate with other resources, and perform data analytics on the gathered information. While possible, extracting residential electrical fire data from fire investigation reports, in the current format, would be a labor-intensive process.

4.1.3.3 *Gaps and Challenges in the Data and Data Analytics*

One of the key challenges associated with using fire investigations as a source of residential electrical fire data is the significant amount of time it takes to conduct investigations and gather enough information to make reliable assessments. The limited sample size is an additional limiting factor for using fire investigation reports to analyze the residential electrical fire problem. The format of fire investigation have to be assessed independently, which can become a cumbersome effort as the quantity of needed investigation reports for such an analysis.

Despite these challenges, fire investigation reports provide the most granular detail regarding electrical fire incidents among the traditional datasets evaluated. Fire investigation reports are also capable of collecting all data elements required for a retrospective approach (e.g. date of construction, wiring type, electrical failure, causal factors, protection devices installed, etc.).

4.2 Emerging Datasets

With the evolution of the Internet of Things (IoT), the ways in which data can be collected is constantly changing. In this data-centric society, there are now new ways to collect information that was previously inaccessible. The evolution of database companies, smart cities, and open city data has enabled public, open-access to fire incident data, fire department response data, housing inventory data, building information data, etc. A few examples of emerging datasets with respect to residential electrical data are discussed within this section.

4.2.1 Open Data Portals

Open Datasets have started to emerge over the last few years. According to the Open Data Handbook, "open data is data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share-alike" (Open Knowledge International, 2018) In the United States, 263 out of 19,354 cities are providing open-source data (1.4%). Of the 263 cities with open data, there are 656 available datasets – only 55% of which are open-access (Open Knowledge International, 2018). Although open datasets are just starting to emerge, there is great value in the currently available data retrieved from open data portals.

4.2.1.1 Positive Characteristics

Accessibility. The most notable characteristic of these datasets is their accessibility. For cities that have open data, a vast array of datasets can be downloaded at "*data.cityname.gov"* and exported into a .csv file. These datasets can be accessed by anyone, at any time.

Format. The format of open-data is uniform and can easily be combined with other datasets to enhance a particular analysis. For instance, datasets such as fire incident data, fire department response data, and building inventory data can be consolidated to enhance the assessment.

Ease of Analysis. The format of open datasets enables them to be analyzed much easier. Since it is in an excel file, the categories of data can be easily filtered to analyze specific portions of the datasets. It also allows other datasets to be easily combined and assessed with respect to other overlaid datasets. Although all information needed is not in the same spreadsheet, they are all in the same format, enabling various datasets to be combined into one spreadsheet. For instance, a fire incident spreadsheet may indicate the type of fire incident, type of unit, and the address of the fire incident, however it likely will not indicate the construction type, age of building, date of renovation, etc. Many cities, however, have building inventory or property assessment data which provides this information about the residential units. While this format still has its limitations and combining datasets can still be time consuming, this digital format still provides significant advantages when compared to the format of the traditional datasets.

4.2.1.2 Negative Characteristics

Incomplete Dataset. Since open data has just started to emerge over the past few years, these datasets are not yet fully developed. However, the concept of open data portals is growing widespread and being adopted throughout the U.S. Although, these datasets are expected to grow and evolve immensely over time, however, the majority of U.S. cities do not currently have open data.

Inconsistency. Since cities are just starting to embrace the concept of open data by writing open data policies and developing open data portals, neither the process nor content is currently standardized. At this time, only select cities have fire incident data and building inventory data that specifies the type of construction, residence classification, year of construction, date of renovations (if applicable), etc. Of the cities that have fire incident data, the quantity and quality of information related to the fire incidents varies significantly.

Lack of Detail. Since the datasets are in an excel spreadsheet, the information contained therein is relatively vague and lacks important details.

4.2.1.3 Gaps and Challenges in the Data and Data Analytics

The primary challenge associated with open data sets is the inconsistency of the data that is currently available within various cities and the lack of detail within the datasets. The cities that currently have robust open datasets are in large urban areas. Small rural towns have little, if any, open data available at this time. It is important to note that open data portals are just starting to be developed and made accessible to the public, these datasets are expected to continue to evolve over time.

4.2.2 Real Estate Websites and Databases

Real estate websites such as Zillow®, Trulia, Realtor.com, LoopNet™, Homes.com, Hotpads™, Frontdoor.com, Yahoo Homes, etc. have emerged over the past decade or so. These websites are able to collect large quantities of information about homes from data sources including broker listing feeds, user-submitted information, county records and other sources (Zillow Group, 2018). These real estate websites capture data on over 90% of the homes in the Unites States. A representative sample of the data elements collected by many of these resources on residential homes are listed in the table below.

Table 1 Residential Data Elements from Real Estate Websites and Databases

Representative Sample: Type of Data Collected on Residential Homes		
Number of Beds	Year Built	Exterior Material
Number of Baths	Number of Stories	Heating Type
Square Footage of Home	Construction Material	Roof Type
Square Footage of Property (lot)	Foundation Materials	Room Count
Type of home (i.e. single/multi-family)	Address(Street, City, State, Zip)	Structure Type

4.2.2.1 Positive Characteristics

Large and Evolving Dataset. These web-based databases collect a large quantity of data on homes within the United States in a relatively consistent format. According to the U.S. Census Bureau, there were approximately 126.22 million households in the U.S. in 2017. For example, Zillow collects data on approximately 91% of the residential homes in the U.S (Zillow Group, 2018).

Important Property Details. Property details such as date of construction, date of renovation, type of home, construction materials, structure type, etc. are provided.

Indication of Fire Damage. If there is fire damage, it is sometimes indicated in the description of the home on the respective webpage (these homes are often marketed as “fixer-uppers”). Though this information can be somewhat limited.

Constantly Updated. Real estate websites are continuously assessing their datasets, updating the information contained in them, and acquiring new information. Thus, it can be assumed that the information stays relevant.

4.2.2.2 Negative Characteristics

Lack of Historical Data. At this time, most real estate websites and databases do not collect historical data (i.e. homes that have been destroyed and/or demolished). Data on residences damaged and/or destroyed by fire is generally not collected at this time.

Variance in Data Quantity. This data is collected through three primary sources: broker listing feeds, user-submitted information, and county records. County records exist for all homes, but this information is very generic and vague. Broker listing feeds is general information about the home, which is useful, however, this is only available if the home is listed. User-submitted information usually provides the greatest detail, however, whether information is submitted and the respective level of granularity is at the discretion of the user.

Lack of Fire Incident Details. If a fire incident is indicated on the listing page for a particular home, generally, little to no detail is provided regarding the fire incident within the home.

Validity of Data. While the majority of the data from real estate databases is accurate, missing information may be approximated in order to make estimates on the value of the property. Therefore, the source of the data should be validated before using it.

4.2.2.3 Gaps and Challenges in the Data and Data Analytics

Currently, the primary focus of real estate data is on the residential market trends (e.g. what is the estimated value of the home compared to last year, is it a buyer or seller's market, etc.). And as a result, it does not track historical data of damaged or demolished homes.

Large loss residential fires typically result in the home being demolished and the damage from small loss fires is generally fixed prior to putting the house back on the market. However, a home will occasionally be listed with fire damage with the intent is to target a renovator to purchase the home at a low price. Despite the indication of fire damage, specific data elements relevant to this analysis such as wiring methods, casual factors of the fire, type of electrical failure, type of branch circuit protection devices installed, etc. are not provided within these databases. However, some of valuable information such as the date of construction, date of renovation, type of residential occupancy, and type of construction could be matched with other fire incident data to assess specific fire incidents.

4.2.3 CPSC Safer Products Database

The consumer products safety commission (CPSC) hosts a searchable database on saferproducts.gov which allows consumers or businesses to report incidents, failures, recalls, etc. on various products (CPSC, 2018). The incident report allows the following information to be collected:

Table 2 Contents of a CPSC SaferProducts.gov Incident Report

Report Details	Product Details	Incident Details	Victims Involved	Additional Details
Report No.	Product Description	Incident Description	Injury Information	Does the submitter have the product?
Report Date	Product Category	Incident Date	Relationship to the victim	Product was damaged before incident?
Sent to Manufacturer /Importer?	Product Type	Incident Location	Victim Gender	Product was modified before incident?
Report publication date	Product Code		Age of victim at time of incident	Have you contacted the manufacturer?
Category of Submitter	Manufacturer/ Importer Name			
	Brand Name			
	Model Name/No.			
	Serial Number			
	Date Manufactured			
	Retailer and Location			
	Purchase Date			

This resource provides a unique way to gather focused information specific to successes and failures of AFCI devices.

4.2.3.1 Positive Characteristics

<u>Information on AFCI Effectiveness.</u> The greatest benefit of the CPSC saferproducts.gov database is that it provides specific details regarding the success or failure of an AFCI during specific incidents. This characteristic is difficult and rare to find in other databases.

<u>Device Specific Information</u>. When reporting an incident to this database, you must indicate the manufacturer of the product, serial number, etc. This type of information allows the incident to be evaluated from the standpoint of why the device failed (e.g. was there a recall on this specific product that contributed to it not operating properly?).

<u>No Geographic Limitations.</u> Any consumer throughout the United States can report an incident into the CPSC SafeProducts.gov database. Therefore, there are no geographic limitations in terms of the data collected. Incidents reported are in both large cities and small rural towns and the residential occupancies may vary between one or two family dwellings to residential high-rises.

4.2.3.2 Negative Characteristics

<u>Limited Number of Reports</u>. Since this is a consumer-based incident submission database, the number of reports is limited to the number of consumers that report incidents. The minimal data provided by this data source could be an inhibitive factor to using this database as a resource.

<u>Reliability of Submitted Information.</u> The information submitted into this database is not required to be verified or vetted for technical validity. Therefore, there is significant uncertainty around the accuracy and reliability of the consumer-submitted data.

<u>Inconsistency; Dependent on Consumer Submissions.</u> The quantity and quality of the reports are dependent on the input of the consumer. Some consumers write incident reports with extensive detail and photographs, while others provide minimal information and leave many fields blank. This database is entirely reliant upon consumer submissions, thus the quality and quantity of reports is inconsistent and the accuracy is sometimes questionable.

<u>Limited information.</u> All information needed for a retrospective analysis of residential electrical fire incidents cannot be provided through the CPSC database.

4.2.3.3 Gaps and Challenges in the Data and Data Analytics

At this time the data provided through the CPSC safeproducts.gov database is extremely limited in quantity. Although this resource often fails to provide in-depth information regarding the fire incident, there are benefits to using this database as a supplemental reference even though it has an insufficient quantity of data to perform in-depth data analytics on the performance of branch circuit protection devices.

4.2.4 Building Information Modelling (BIM)

Building information modeling (BIM) is "a process involving the generation and management of digital representations of physical and functional characteristics within buildings" (National Institute of Building Sciences, 2015). BIM is primarily utilized during the engineering and construction phase of developments, however, BIM is now starting to be used as a data resource for documentation and on-going maintenance.

While the focus of the BIM technology started with structural systems, it has now expanded to other building systems such as HVAC, mechanical, electrical, and plumbing (Lack, 2016). Electrical manufacturers are starting to provide BIM friendly data for various manufactured electrical products to integrate these products into the building system. Adding and integrating more information into the model at the onset of a project improves the long-term value of the model and provides significant value to on-going building management and maintenance efforts.

The characteristics of this resource are described in sections 4.2.4.1 and 4.2.4.2 below.

4.2.4.1 Positive Characteristics

Detailed documentation of building materials and products. As the building is constructed, all building materials, wiring types, electrical products installed, etc. will be documented through BIM management systems.

Real-time updates. Any change to any aspect of the building will be documented in real-time. Therefore, the data extracted from BIM is ensured to be up to date.

Data Management. BIM also enables data management. Beyond the information that is visually represented, it collects and maintains data on product specifications and performance data (Lack, 2016).

4.2.4.2 Negative Characteristics

Lack of widespread use. Although building information modeling has been around for a significant period of time, its use and implementation is just starting to pick up. While BIM is now being used on a much broader scale, most existing buildings will not have used BIM for the design, construction, and maintenance of their buildings. This will only be a means of collecting data for current and future buildings – and will still be restricted to those buildings which are designed using this platform.

Few BIM Datasets. Building information modeling is now starting to be used for architects, engineers, and contractors, however, most existing buildings did not utilize the BIM software, so the number of building information datasets is limited – particularly for residential properties.

Limited use for residential projects. At this time, BIM is more commonly used for commercial projects rather than residential – therefore, the available residential building data from BIM datasets is likely fairly limited, if any at all.

Data Access. Another limiting factor is getting access to the data. The BIM data is currently accessible by the architects, engineers, contractors, and facility managers that contributed to the design or are involved in the data management piece. This data is not open-access at this time.

Determining Use. How to determine whether BIM was used on a residence which has encountered an electrical-related fire is still relatively unknown.

4.2.4.3 Gaps and Challenges in the Data and Data Analytics

Currently, there are challenges utilizing BIM as a source of building data. These challenges are primarily centered on its limited use for residential occupancies and the restricted access to the BIM data for anyone beyond the design team. BIM data does not contain any information on fire incidents, casual factors, or failure types because the model merely documents the building features and products contained within. Therefore, BIM data, if accessible, can only be used as a supplemental resource to the other data sources used to evaluate residential electrical fires and the effectiveness of AFCI's. The data available through BIM models would essentially be used to document the electrical features of the building and the date of construction.

4.2.5 Others

It is acknowledged that the emerging datasets discussed throughout this section are not intended to be all-encompassing. However, the datasets mentioned herein are common platforms for data acquisition on residential fire incidents, housing and infrastructure data, electrical information, product information, etc.

4.3 Summary Observations for the Existing Data Landscape

As discussed in Sections 4.1 and 4.2, data collection on residential electrical fires within the existing landscape is generally independent and incomplete. While proving the effectiveness of preventative measures (e.g. AFCI's) is a challenging task, the significant limitations associated with the existing traditional data sources presents serious concerns. An additional challenge with the existing data, is that the data elements needed are siloed into distinct categories of data (i.e. building data, fire incident data, electrical data, and code enforcement data). The independent (siloed) and incomplete data places significant constraints on the ability to perform data analytics through a retrospective, incident centric approach.

A unified database with the integral building data elements, fire incident data elements, electrical data, and code enforcement data is lacking within the existing data landscape – prohibiting a holistic view of the residential electrical fire problem. As noted in Figure 8 below, an interconnection of residential electrical data elements to create a unified residential electrical database is needed.

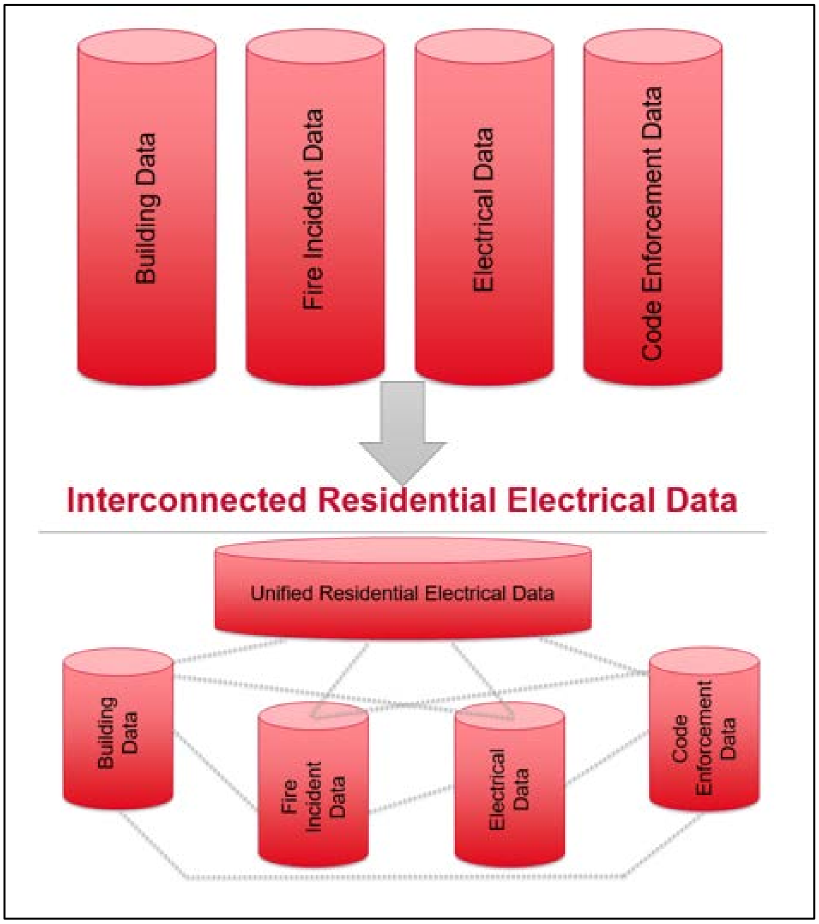

Figure 8 Need for Interconnection of Residential Electrical Data

All necessary data elements can in some way be collected from the currently available and emerging data sources, however, all data sources have their shortcomings. Table 3 below evaluates the attributes and shortcomings of some of the existing data sources. These characteristics include: data quality, reliability of content, granular detail, data quantity, data accessibility and compatibility, data format, relevancy, etc.

Table 3 Assessment of Data Source Characteristics for Residential Electrical Fire Data

	Adequate Quality	Reliable Content	Granular Detail	Adequate Quantity	Accessible	Ideal Format	Relevant Data
Traditional Datasets							
NFIRS	No	Moderate	No	Yes	Yes	No	Moderate
FIDO	Yes	Yes	Yes	No	Moderate	No	Moderate
Fire Inv.	Yes	Yes	Yes	No	Moderate	No	Moderate
Emerging Datasets							
Open Data Portals	No	Yes	No	No	Yes	Yes	Moderate
Real Estate Databases	Yes	Yes	No	Yes	Yes	Moderate	Yes
CPSC	No	No	No	No	Yes	No	Yes
BIM	Yes	Yes	No	No	Moderate	Yes	Yes

Note: This illustration is entirely focused on these datasets with respect to residential electrical fire data for collection of retrospective data. It should be noted that other categories of data within the datasets mentioned herein may have different characteristics.

As demonstrated in Table 3, none of the existing databases have all of the characteristics desired for residential electrical fire data. Thus it will be important to leverage the use of the internet of things, cyber physical systems, wireless sensor technology, artificial intelligence algorithms, smart grid applications, etc. to inform the evaluation of the residential electrical fire problem through future efforts.

The following section addresses the current trends in the data landscape for residential electrical data and future and anticipated datasets that are being proposed, developed, or adapted to and for electrical applications.

5. Trends and Future Envisioned Datasets

Today we live in a sea of data – it is everywhere and growing exponentially. Although the existing data landscape presents significant challenges from a data analytics standpoint, accessibility to data and novel data acquisition methods is rapidly evolving. In 2012, 2.5 billion GB of data were generated daily, while 1.7 megabytes of data is expected to be generated every second for every person in the world by 2020, according to IBM (Monnappa, 2018). This is largely a result of the evolution of the Internet of Things (IoT).

The Internet of Things, defined as “an interconnected cloud-based network of sensors that communicate through connection to the internet”, is revolutionizing data collection and analysis approaches (Invicta, 2018). The IoT is rapidly accumulating countless sources of data from extensive sensor networks. This technology revolution is making our world increasingly sensor rich. These sensors are in our phones, computers, vehicles, fire and life safety equipment, building systems, electrical systems and devices, etc. The sensors incorporated into various products and infrastructure is a result of the increased computer processing speeds, increased storage capacity, and the reduction in technology costs.

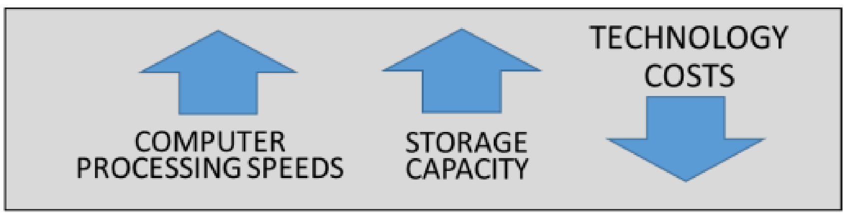

Figure 9 Characteristics to Enable a Sensor Rich World (Grant, C., et. al (2015))

Since all of these sensors are connected to the internet, the recorded data can be uploaded, consolidated, and further analyzed to develop meaningful information. This technology is providing access to new information and more robust datasets that were previously unavailable.

The foundation for the development and evolution of smart devices, smart homes, smart buildings, and smart cities has been built on the backbone of IoT enabled technology. This technology has capabilities of cross-referencing compatible data sources which enables the use of data analytics to optimize the quality of information provided to better inform complex decision-making regarding electrical fire safety.

Now, our computer systems, software, and complex algorithms have advanced to a point that enables customized analyses of “Big Data” that have been collected and will continue to be collected going forward.

With the development of such technologies and the resultant output of data, data scientists are beginning to incorporate artificial intelligence (AI) based algorithms to perform big data analytics on these various sources of data. Examples of data analytic techniques utilized include machine learning, deep learning, data mining, text mining, image recognition, speech recognition, and others.

While future datasets are not expected to collect the exact data points needed for a retrospective study on residential electrical fires, a prospective study is able to take a different approach in terms of the data that needs to be collected. This approach will differ from looking at what happened, to looking at what is happening in the electrical system. The success of a prospective approach will depend on the ability of future datasets to leverage the use of new and emerging technologies to acquire quality and valuable data on residential electrical systems.

How trending concepts of IoT and artificial intelligence are impacting technology and the corresponding collection, analysis and communication of residential electrical data is discussed within Section 5.1.

5.1 IoT and Artificial Intelligence Driven Technology

In recent years there has been significant attention regarding the concept of artificial intelligence. Artificial intelligence (AI) is defined as "the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages" (Marr, 2018). While AI is not a new concept, it was its application to the internet of things (IoT) revolution that the true value of each were realized. Until recently, the data has always been the limiting factor (Burns, 2018). However, the mass quantities of data being produced by IoT devices and others is presenting a huge opportunity to incorporate artificial intelligence based algorithms into emerging technologies.

Machine learning and deep learning are two factions of artificial intelligence which are being used for various applications including recognition of abnormal trends, image recognition, speech recognition, chatbots, natural language generation, sentiment analysis, etc. (Burns, 2018). These are defined below:

Machine Learning – *"Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed"* (Expert System, 2018).

Deep Learning – *"Algorithms attempting to model high level abstractions in data to determine a high level meaning"* (Leone, 2018).

Machine learning, in particular, is having an immense impact on data analytics for electrical applications. The capabilities of these AI-based applications are advancing to the point of training themselves to recognize complex patterns in a stream of data. While their application to electrical systems is relatively new, the use of these advanced algorithms is rapidly expanding. While artificial intelligence driven technology is still evolving, there are a few examples of developed devices that utilize artificial intelligence algorithms to monitor, detect, and report relevant information from the raw data collected on electrical systems. Some examples discussed in Sections 5.1.1, 5.1.2, and 5.1.3 include electrical monitoring and sensing devices, smart power outlets, and the smart grid application, respectively.

5.1.1 Connected Electrical Monitoring Devices and Applications

This section contains three relevant examples of electrical monitoring devices and applications that are currently being developed and can be anticipated in the future. It should be noted that the technologies mentioned herein are still in the development phase and have yet to be fully vetted or technically validated for electrical systems. Further research for their applicability for detecting hazards in electrical systems is needed.

5.1.1.1 Electrical Monitoring and Sensing Devices

As the Internet of Things (IoT) continues to grow and sensors are developed and incorporated into equipment, systems, and devices within buildings, the condition of various systems and their operating efficiency can be monitored. When connected to the electrical system and connected to Wi-Fi, this technology can potentially monitor electrical scintillations flowing through the electrical network in a home and collect electromagnetic data up to nearly 30 million times per second (Jacobson, 2018). With the continual influx of data, machine learning algorithms can be used to analyze the data and may potentially be able to differentiate between normal and abnormal conditions in the electrical system that could result in a fire. As data is continually gathered, the accuracy of the model detecting fire hazards in the electrical system will likely improve over time as the model becomes trained on the meaning of electrical signatures and system behaviors.

Since this technology is an IoT enabled device, its connection to the internet enables the valuable information derived through data analytics to be communicated to the appropriate parties. This line of communication presents opportunities to increase awareness of the health and status of the electrical system. The positive and negative characteristics of this technology are discussed in sections 5.1.2 and 5.1.3.

5.1.1.2 Smart Power Outlets

Another IoT and AI enabled device being developed are smart power outlets. These smart outlets claim to incorporate machine learning algorithms which are able to differentiate between harmless arcs and hazardous arcs that can result in a fire (Brown, 2018). While this technology has similarities to arc fault circuit interrupters, these emerging smart power outlets have significantly more complex algorithms incorporated into the technology to potentially enhance their ability to accurately detect truly hazardous arcs and minimize false activations.

5.1.1.3 Smart Grid

The smart grid is defined as "an electrical grid which includes a variety of operational and energy measures including smart meters, smart appliances, renewable energy resources, and energy efficient resources" (U.S. Department of Energy, 2010). With respect to electrical systems, the smart grid is a platform that leverages the use of cyber-physical systems, smart sensors and AI-based algorithms to advance the metering infrastructure. As a result, there is a mass deployment of smart devices being installed into the infrastructure of generation units, transmission lines, and distribution level systems.

Due to the depth of technology incorporated into the smart grid, there will be massive quantities of data produced. From the nearly 2 million customers currently using the smart grid infrastructure, approximately 22 GB of smart meter data is generated per customer, per day (IEEE Smart Grid Working Group, 2018).

The smart grid is essentially an integrated system of smart sensors with embedded AI-based algorithms that monitor and assess energy usage, operations, current and voltage measurements, etc. The smart grid uses smart meters to collect data on both residential and commercial properties to better inform utility companies of the electricity demand. Although the desired data output for utility companies may be relatively narrow in scope, the sensors continue to monitor the electrical system and gather information on its performance, behavior, usage, etc. While the application to the utility grid differs from the other technologies discussed (e.g. sensing and electrical monitoring equipment and smart power outlets), the concept behind the smart grid technology is similar and collects data relevant to this study. Therefore, it also has similar positive characteristics (e.g. wireless connection, data reporting, continual monitoring, and data processing) and negative characteristics (e.g. data accuracy in the beginning, data privacy, data access), while also providing similar benefits (optimal data quantity, quality, and format) with respect to data analytics.

5.1.2 Positive Characteristics

The positive characteristics listed below are generally applicable to the three emerging technologies/applications mentioned in the previous sections (5.1.1.1, 5.1.1.2 and.5.1.1.3).

Large Dataset. Since such large quantities of data are collected every second, every device connected to an electrical system has the ability to produce large quantities of raw data to be analyzed for each home. This large dataset could provide an adequate quantity of data to evaluate the residential electrical fire problem.

Reliable Data. The algorithms incorporated into these devices have been trained to detect a specific combination of signatures as an electrical arc before a visible fire is ignited. The reliability of this data is highlighted by the scientific basis to detection of such conditions.

Wireless Connection. These smart devices are integrated as part of the internet of things, which also has the ability to connect wirelessly and communicate with other devices (Brown, 2018).

Timely Data Reporting. The data reported can include electrical usage, appliances in use, and notification of a tripped device and possibly even the respective cause. Since data is continually collected on the electrical system, the status of the electrical system will likely be able to be viewed in real time, presenting minimal delays in receiving notification of a hazardous condition.

Continual, Real Time Monitoring. Sensors can continually take readings from several thousand to several million times per second; thus the reporting would always be relevant. Since these devices contain a memory element, they are able to monitor the activity of the electrical circuit in real time.

Processes Data. Because these technologies have machine learning algorithms integrated into the devices, the embedded software is able to analyze the data as it is collected to report meaningful information (e.g. what device tripped). With the advanced machine learning algorithms integrated into these devices, these technologies can be trained to take the appropriate action to mitigate the hazard (e.g. shut down the electrical circuit), when a hazardous condition is detected.

5.1.3 Negative Characteristics

The negative characteristics listed below are generally applicable to the three emerging technologies/applications mentioned in the previous sections (5.1.1.1, 5.1.1.2, and.5.1.1.3).

Still in development. The technologies mentioned in sections 5.1.1.1, 5.1.1.2, and 5.1.1.3 are not all currently available in the marketplace. They are still generally in the development phase. It should be noted that since these technologies are primarily still in the development phase, they have not all been fully vetted or technically validated for electrical systems. Further research for their applicability for detecting hazards in electrical systems is needed.

Undue endorsement. Since these technologies are not fully developed or deployed, the endorsement of such technologies and their respective abilities is not currently justified by the electrical industry. The technologies and applications mentioned herein need to be technically validated by the electrical industry prior to utilizing these technologies as a means of data collection for residential electrical systems.

Data accuracy evolves over time. The accuracy of the detection of arc faults may not be perfected in the beginning. As more data is collected, the accuracy of autonomously being able to detect electrical arc faults will improve.

Data Access. There is significant uncertainty around the concept of data access in such “smart devices.” Whether the data is owned by the owner of the home or occupancy, the electric company, or the product developer is unknown and could potentially be a concern with utilizing this data.

5.1.4 Data and Data Analytics

Due to the format and quantity of data collected through these technologies, there is great potential for valuable analysis. Since the embedded sensors are continually monitoring the electrical system, there will be a large quantity of data available to enable realistic conclusions to be drawn from the data and implemented analytic techniques. The advantage of applying machine learning or incorporating other AI-based algorithms into devices is that the data analytics is already embedded into the technology (Wiggers, 2018). These technologies and advanced algorithms are taking the raw data, processing it, and delivers it to an app or software which reports the data as valuable, processed information to the end-user.

The challenge will be getting access to the data and finding solutions to concerns of data privacy and security to permits analytics to be performed on the dataset. Ownership of sensor data from residential homes is still a broadly unresolved issue.

5.1.5 Others

It is acknowledged that the future technologies discussed throughout this section is not an exclusive list, as there are a vast array of technologies available or under development related to this focus area. These technologies mentioned herein are examples of common platforms under development for data acquisition of information on residential electrical systems.

While the technologies mentioned in 5.1.1 are not the only technologies being developed on which residential electrical data can be collected, the data sources mentioned herein provide insights into trends of future data collection for residential electrical systems. The technologies being developed are commonly IoT devices that incorporate advanced algorithms to enable “big data” to be consumed and processed to learn from the common behaviors and enhance the ability to detect hazardous conditions with greater accuracy.

5.2 Challenges with Future Datasets

These future datasets and technologies present significant advancements in how the residential electrical fire problem can be assessed. In fact, many of the desired dataset characteristics including data quality, data quantity, data format, relevancy, reliability, and others are able to be met with these future datasets. However, since these technologies are still relatively new to the market, with many still in development, there are still an array of issues that have yet to be resolved. Examples of various issues that need to be addressed are listed below:

Missing Data. When a sensor goes bad and the data is not able to be collected, there may be a period of incomplete or missing data. The complex algorithms incorporated into these sensor technologies are extremely reliant on the collection of good quality data, since the model continually learns from the available data. How to deal with periods of "unavailable" data is still relatively unknown (IEEE Smart Grid Working Group, 2018).

Cost. To be able to collect data from sensors, these devices either need to be incorporated into existing equipment, or the existing equipment needs to be replaced with equipment that has AI-based sensor technologies incorporated into them. The replacement or upgrade of the existing infrastructure could require a significant financial undertaking which could restrict the use of these technologies.

Data Volume. How to handle the massive quantities of data that will continually be collected via sensors is not fully resolved. While data storage in the "cloud" is generally capable of handling large volumes of data, the scale of data that these technologies may be producing presents concerns that should be addressed further.

Inadequate Training Data. While the machine learning has drastically advanced over the past few years, its ability to detect hazardous fire conditions in an electrical system is reliant on the quality, quantity, and accuracy of the training data. Inadequate training data can reduce the confidence and reliability of the assessed data (IEEE Smart Grid Working Group, 2018). Thus, it is critically important to ensure the training data is accurate and of adequate quality – this should be confirmed by electrical experts.

Data Privacy and Security. Issues of data privacy, security, and ownership have become important questions in our evolving data-rich world. Questions such as "Who owns the data? Who needs what data? Who has access to what data? Who has the rights to provide access to others?" do not have clear answers.

Technology Lacks Credible Validation for Electrical Applications. The ability and accuracy of these technologies identifying hazardous conditions within electrical systems is not technically validated by credible electrical experts and needs further research and analysis prior to full implementation for industry use. The endorsement of these technologies is not fully justified at this time.

6. Summary Observations

Data collection is the fundamental first step to evaluating the effectiveness of electrical branch circuit protection methods. The focus of this report has been on data collection for residential electrical systems and fire incidents – outlining what data sources are currently available, emerging, and anticipated in the near future.

The ultimate goal is to determine how to collect data on residential electrical systems and respective fire incidents that have the desired quality, granularity, quantity, and format needed to enable valuable insights regarding the performance of arc-fault detection devices to be developed. This can be achieved through either a retrospective or prospective study. However, the data needed for each approach varies significantly, as discussed in Section 3.

Data of adequate quality, quantity, and format that can be coordinated with various data sources is lacking in the present infrastructure. The vast variations in data characteristics and formats hinders the ability to analyze and develop insights from past residential electrical fire incident data. As our world transitions to a more digital landscape of data collection, coordination between various data repositories is essential. Thus, as future equipment, technologies, software, and databases are developed, it is important to ensure specific performance attributes such as data availability, reliability, operability, maintainability, and compatibility are incorporated and maintained in the emerging and future data landscape for residential electrical systems.

Currently, the existing residential electrical data is generally unrefined and provides limited value to the analysis of determining the effectiveness of electrical branch circuit protection devices. Data is currently collected in a series of independent silos, but the different sectors of relevant data are generally not compatible or given context in relation to each other. There is a distinct need for data sources to be interconnected. Interoperable data could potentially promote communication between data sources, visibility of system performance, and provide a better understanding of the residential electrical fire problem. The future vision should focus on collecting large quantities of good and accurate data that is compatible, scalable, and unified to create “Big Data” for residential electrical systems as shown in Figure 10 below. As more data is collected, the knowledgebase around residential electrical fires will expand, valuable insights may be developed, and the ability to make more informed decisions regarding the most effective way to protect electrical branch circuits against electrical arcing could be improved.

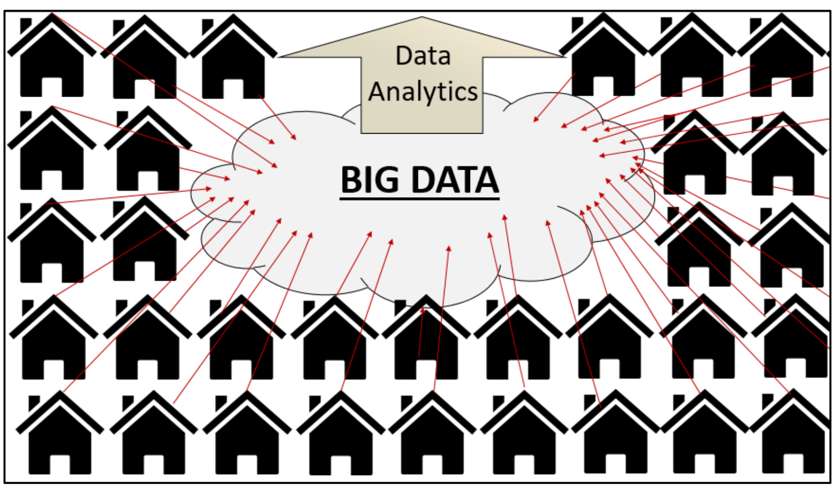

Figure 10 Concept of Big Data for Residential Electrical Systems

Per the assessment of the existing, emerging, and future data landscapes, key observations and recommendations for continuing to address the gaps and challenges associated with residential electrical data are provided below.

1) **<u>Good and accurate</u> data is needed**. The most significant problem with residential electrical fire data is that nearly all of the currently available public data is lacking in quality and accuracy, and is relatively unusable for data analytics in its current state. If the goal is to utilize data to evaluate the effectiveness of branch circuit protection devices, it is imperative that we collect better data. Without better data, we cannot prove the effectiveness of these devices or evaluate the optimal means of protecting branch circuits against electrical arcing.

   To evaluate the effectiveness of preventative devices, such as AFCI's, data is needed that can show a hazardous condition in the electrical system and a device activating to mitigate the hazard. Though it is understood how to use the data, the analysis cannot be performed without first acquiring quality data.

2) **<u>A lot</u> of good data is needed**. In addition to quality data, we need a high volume of data. Within this large quantity of data, it should also be from a cross section of residential applications (e.g. residences in both large urban cities and small rural towns; single family dwelling and a residential high-rise, etc.). The primary issue with getting good data within the existing infrastructure is the frequent misclassification of electrical fire incidents, which limits the quantity of good, available data. Further challenging this analysis from a data perspective is that we need thousands of data points from fire incidents to adequately evaluate the residential fire problem with respect to electrical arcing.

   While past data is difficult to get in such large quantities, new technologies mentioned within this report are producing data at rates of several million readings per second. Therefore, in future datasets, the desired quantities may be available, however, the limiting factor will be the diversity of the dataset and the availability of installed devices to collect data from.

3) **Data analytics <u>is extremely challenging</u> to perform on existing post-incident data.** The quality and quantity of currently available residential electrical fire incident data is inadequate. Much of the necessary data is missing, incomplete, or in an unusable format. To analyze the data collected via existing datasets, all data elements will have to be mapped to a particular incidents, then combine all the data with other incidents collected, and manually analyze the data. This is a challenging effort. In addition, the quantity of good data collected via existing, incident centric datasets further restricts the ability to perform analytics on the collected data.

   However, the data produced from current technologies and devices are being collected in a way that is optimal for performing data analytics. These emerging technologies are designed to provide analytics on the collected data, as advanced algorithms are incorporated into the technology itself. Thus, despite the data analytic challenges with existing datasets, there is great potential for data analytics to produce valuable information and knowledge from emerging and future technologies and databases.

4) **The data needs to be <u>compatible, unified, and scalable</u>.** While data can be collected through various data sources, mining and evaluating the information is the challenging piece. Compatibility issues between diverse datasets limit utilization of the data. Having the data in a compatible format that supports unification of data sources to leverage valuable insights is critical, going forward.

5) **The greatest challenge lies with the <u>non-technical issues</u>.** In today's society we are rapidly developing solutions to technological challenges. However, the technological solutions are not necessarily the most difficult part. The primary challenges lie with the non-technical legal (e.g. privacy of data, confidentiality, or proprietary information), political, social, cultural, and economic issues that extend beyond the technical challenges – particularly with the emerging and future datasets and technologies. There are a series of unanswered questions which need to be addressed, including: Who owns the data? Can private information be extracted from the data? Can proprietary information be protected to ensure client privacy while not hindering data analytics? Etc.

6) **All data collected, despite the source, needs to be <u>technically validated</u> prior to use.** In order for the data collected to be used, it needs be scientifically vetted and verified that all collected information is factual and accurate prior to being incorporated into a residential electrical data analysis.

The information contained herein summarizes the data landscape for residential electrical systems and related fire incidents. As discussed throughout this report, determining the most effective means of protecting electrical branch circuits against electrical arcing requires collection, processing, and delivery of a lot of good quality, validated data. Before conclusions can be drawn regarding the effectiveness of electrical protection devices, the summary observations listed above should be addressed with respect to residential electrical data.

7. References

Brown, M. (2018, June 20). *No more "Nuisance Trips" with Smart Power Outlets for Smart Homes*. Retrieved from engineering.com: https://www.engineering.com/DesignerEdge/DesignerEdgeArticles/ArticleID/17149/No-More-Nuisance-Trips-with-Smart-Power-Outlets-for-Smart-Homes.aspx

Burns, E. (2018). *AI functionality limited today but could be a game-changer*. Retrieved from TechTarget: https://searchenterpriseai.techtarget.com/feature/AI-functionality-limited-today-but-could-be-a-game-changer

Campbell, R. (2017). *Electrical Fires.* Quincy: National Fire Protection Association.

CPSC. (2018). *SaferProducts.gov*. Retrieved from https://www.saferproducts.gov/Default.aspx

Expert System. (2018). *What is Machine Learning? A definition*. Retrieved from Expert System: https://www.expertsystem.com/machine-learning-definition/

Grant, C., Hamins, A., Bryner, N., Jones, A., & Koepke, G. (2015). *Research Roadmap for Smart Firefighting.* Gaithersburg: National Institute of Standards and Technology .

IEEE Smart Grid Working Group. (2018). *Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid: Introduction, Benefits, Challenges and Issues.* IEEE.

Invicta. (2018). *How the Internet of Things (IoT) is Changing Fire Safety*. Retrieved from The Invicta Group: https://www.theinvictagroup.co.uk/news/internet-of-things-iot-changing-fire-safety

Jacobson, J. (2018, May 30). *Whisker Labs' AI-Driven Ting Could be Huge Breakthrough in Fire Prevention*. Retrieved from CEPro: https://www.cepro.com/article/whisker_labs_ai_ting_iot_home_fire_prevention

Lack, B. (2016, July 6). An Essential Guide to BIM for Electrical Contractors. Retrieved from Schneider Electric.

Leone, G. (2018). *The Best Explanation: Machine Learning vs Deep Learning*. Retrieved from KAIROS: https://medium.com/lovekairos/the-best-explanation-machine-learning-vs-deep-learning-d5c123405b11

Marr, B. (2018, February 14). *The Key Definitions Of Artificial Intelligence (AI) That Explain Its Importance*. Retrieved from Forbes: https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#43aae2164f5d

Monnappa, A. (2018, August 13). *Data Science vs. Big Data vs Data Analytics*. Retrieved from SimpliLearn: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

National Institute of Building Sciences. (2015). *National BIM Standard - United States*. Retrieved from buildingSMARTalliance: http://www.nationalbimstandard.org/files/NBIMS-US_FactSheet_2015.pdf

Open Knowledge International. (2018). *Open Data Handbook*. Retrieved from Open Data Handbook: http://opendatahandbook.org/

Open Knowledge International. (2018). *U.S City Open Data Census*. Retrieved from Open Data Census: http://us-cities.survey.okfn.org/

Romano, J. (2006, February 19). The Fire Dangers of Aluminum Wiring. *The New York Times*.

Thomas, D. S., & Butry, D. T. (2011). *Tracking the National Fire Problem: The Data Behind the Statistics.* Gaithersburg: U.S. Department of Commerce National Institute of Standards and Technology.

U.S. Department of Energy. (2010). *Data Access and Privacy Issues Related to Smart Grid Technologies.* Washington D.C.: Department of Energy: United States of America.

U.S. Fire Administration. (2018). *National Fire Incident Reporting System*. Retrieved from UFSA: https://www.usfa.fema.gov/data/nfirs/index.html

Wiggers, K. (2018, June 15). *MIT Researchers develop a smart power outlet that can prevent your house from burning down*. Retrieved from Venture Beat: https://venturebeat.com/2018/06/15/mit-researchers-develop-a-smart-power-outlet-that-can-prevent-your-house-from-burning-down/

Zillow Group. (2018). *Zillow*. Retrieved from www.zillow.com