

CONFERENCE OF EUROPEAN STATISTICIANS

**Workshop on Statistical Data Collection**

WP.1-1

10-12 October 2017, Ottawa, Canada

16 August 2017

---

## **Evaluating the Use of Publicly Available Building Permit Data to Complement Data Collection and Analysis of Construction Surveys**

Angela Delano (U.S. Census Bureau, United States of America)

*angela.lynn.delano@census.gov*

### ***Abstract***

Government agencies, policy analysts, financial institutions, and manufacturers in the United States use new construction data collected by the U.S. Census Bureau to help measure and evaluate size, composition, and change occurring within the U.S. housing and construction sectors. To measure new construction, the U.S. Census Bureau conducts two construction surveys using data on building permits – the Building Permit Survey (BPS) and the Survey of Construction (SOC).

To one extent or another, the BPS and the SOC represent a nationally representative cross section of all permit-issuing jurisdictions in the United States. These are sample surveys conducted in predefined permit-issuing jurisdictions requiring a building permit to authorize new construction. As with many economic surveys, collection of building permit data requires time and cost commitments that are constantly increasing. Factor in observed decreases in survey response rates and the prospect of expending more resources in an effort to obtain less data is becoming more and more a survey reality. This is especially true for jurisdictions receiving requests from each of the two construction surveys.

In the past several years, permit-issuing jurisdictions have begun making their permit information available publicly. While the content, format, and accessibility of this information varies widely across jurisdictions, it nonetheless makes sense to explore their viability as new data source in more detail. As jurisdictions move further away from traditional survey cooperation, this new survey reality may become more the rule than the exception. This research serves just that purpose, to evaluate the use of publicly available building permit data to complement both data collection and analysis methods for the BPS and the SOC. Lessons learned from this research can then be applied to additional data source formats moving forward.

# Evaluating the Use of Publicly Available Building Permit Data to Complement Data Collection and Analysis of Construction Surveys

Angela Delano

U. S. Census Bureau, Washington, DC; [angela.lynn.delano@census.gov](mailto:angela.lynn.delano@census.gov)

## **Abstract:**

Government agencies, policy analysts, financial institutions, and manufacturers in the United States use new construction data collected by the U.S. Census Bureau to help measure and evaluate size, composition, and change occurring within the U.S. housing and construction sectors. To measure new construction, the U.S. Census Bureau conducts two construction surveys using data on building permits – the Building Permit Survey (BPS) and the Survey of Construction (SOC).

To one extent or another, the BPS and the SOC represent a nationally representative cross section of all permit-issuing jurisdictions in the United States. These are sample surveys conducted in predefined permit-issuing jurisdictions requiring a building permit to authorize new construction. As with many economic surveys, collection of building permit data requires time and cost commitments that are constantly increasing. Factor in observed decreases in survey response rates and the prospect of expending more resources in an effort to obtain less data is becoming more and more a survey reality. This is especially true for jurisdictions receiving requests from each of the two construction surveys.

In the past several years, permit-issuing jurisdictions have begun making their permit information available publicly. While the content, format, and accessibility of this information varies widely across jurisdictions, it nonetheless makes sense to explore their viability as new data source in more detail. As jurisdictions move further away from traditional survey cooperation, this new survey reality may become more the rule than the exception. This research serves just that purpose, to evaluate the use of publicly available building permit data to complement both data collection and analysis methods for the BPS and the SOC. Lessons learned from this research can then be applied to additional data source formats moving forward.

## **1 Introduction**

Government agencies, policy analysts, financial institutions, and manufacturers in the United States use new construction data collected by the U.S. Census Bureau to help measure and evaluate size, composition, and change occurring within the U.S. housing and construction sectors. To measure new construction, the U.S. Census Bureau conducts two construction surveys using data on building permits – the Building Permit Survey (BPS) and the Survey of Construction (SOC).

*Disclaimer: This paper is released to inform interested parties of (ongoing) research and to encourage discussion (of work in progress). Any views expressed on (statistical, methodological, technical, or operational) issues are those of the author(s) and not necessarily those of the U.S. Census Bureau.*

To one extent or another, the BPS and the SOC represent a cross section of all permit-issuing jurisdictions (referred to here as places) in the United States. These are sample surveys conducted in predefined permit-issuing places requiring a building permit to authorize new construction. Surveying of these predefined places takes place on a monthly or annual basis via contact with a local Building Permit Office.

## **2 New Construction Surveys**

### **2.1 Building Permit Survey**

The BPS is one of the most comprehensive new, privately-owned, residential housing surveys conducted in the U.S. The survey provides national, regional, state, and local data. Topics covered by the survey include the number of residential buildings and number of housing units authorized by building permits by size of structure. Permit valuation data by size of structure is also collected. The BPS is funded and conducted by the U.S. Census Bureau.

The BPS is a building permit survey conducted on a continuous basis. Building permit data for the survey is collected from predefined permit-issuing places. Places selected for the BPS represent a cross section of all permit-issuing places in the U.S. Surveying of these predetermined permit-issuing places takes place on a monthly or annual basis. Reselection of new predetermined permit-issuing places takes place approximately every 10 years.

### **2.2 Survey of Construction**

Similar to the BPS, the SOC is also a comprehensive new, privately-owned, residential housing survey conducted in the U.S. The survey provides national and regional data on a range of topics. Topics covered by the survey include construction start and completion dates, construction characteristics (such as number of bedrooms, bathrooms, and square footage, etc.), and sales data (such as sales date, pricing, and financing). Start, completion and characteristic data are provided for both single-family and multifamily housing. Sales data are provided only for single-family housing. The SOC is partially funded by the Department of Housing and Urban Development and conducted by the U.S. Census Bureau.

The SOC is a housing unit survey conducted on a continuous basis. Housing unit data for the survey is collected from predefined permit-issuing places and predefined non permit-issuing areas. Permit-issuing places selected for SOC are, by design, also monthly reporters to the BPS. Housing units selected for the SOC represent a cross section of all new, privately-owned, residential housing units constructed in the U.S. Reselection of new predetermined permit-issuing places takes place approximately every 20 years.

Each month a sample of housing units is selected for survey interview through one of two methods. The first method includes sampling of building permits in predetermined places requiring a building permit to authorize new housing unit construction. Surveying of these predetermined permit-issuing places takes place on a monthly basis. The second method includes physical canvassing of predetermined areas that do not require a building permit to authorize new housing unit construction. Surveying of these predetermined non permit-issuing areas takes place on a monthly or quarterly basis.

Once a housing unit is selected, the survey returns to the same housing unit approximately every month from authorization of construction (or discovery via canvassing) through completion (or occupancy) and, if necessary, sale (for single-family housing). Respondents for the survey consist of builders, owners, and sales agents of new housing units.

The BPS and the SOC are sample surveys. Estimates based on sample surveys may differ from estimates obtained from a complete census. The accuracy of an estimate based on a sample survey is subject to the joint effects of both sampling errors and nonsampling errors. For the purpose of this research, we will not discuss these errors. See the Reference section for additional information regarding these surveys.

### **3 Data Collection Issues**

#### **3.1 Duplicative Data Requests**

Research by Hak, Willimack & Anderson (2003) found respondent reporting burdens associated with duplication of effort in providing similar information to multiple survey requests. This is especially true of permit-issuing places that may receive requests for BPS and SOC data during the same survey month.

#### **3.2 Survey Nonresponse**

Issues regarding data collection for these surveys are similar to those of establishment surveys. Sudman et al (2000) reported data collection for establishment surveys entails specific respondent processing and reporting burdens that may negatively affect data quality. In addition to these issues, rising survey costs and falling response rates must be factored. Tourangeau & Plewes (2013) found response rate problems in not just government surveys, but other surveys as well.

Table 1 on the following page shows reporting source distributions for the monthly BPS from January 2016 to January 2017. Over this timeframe an average 33% of permit-issuing places reported no timely residential data. In these situations, if there is no additional data source available (i.e. SOC), missing data will be imputed for the survey, contributing to data quality issues.

Reporting Source	Monthly Building Permit Survey - January 2016 to January 2017												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
<i>Reports Received</i>	68%	69%	68%	65%	66%	66%	67%	65%	66%	67%	67%	69%	68%
Paper Reporting	40%	41%	41%	40%	39%	39%	39%	36%	37%	37%	38%	39%	38%
Electronic Reporting	4%	4%	4%	3%	4%	4%	4%	2%	3%	3%	2%	2%	2%
Internet Reporting	24%	24%	23%	22%	23%	23%	24%	27%	26%	26%	27%	28%	28%
<i>Reports not Received</i>	32%	31%	32%	35%	34%	34%	33%	35%	34%	33%	33%	31%	32%
Data from SOC	4%	4%	4%	4%	4%	4%	4%	4%	4%	4%	4%	4%	3%
Data from Imputation	28%	27%	28%	30%	30%	29%	29%	30%	29%	29%	28%	26%	28%
Data not Received	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
<i>Sample Size</i>	8,457	8,457	8,460	8,457	8,457	8,456	8,454	8,454	8,454	8,455	8,453	8,454	8,455

**Table 1:** Reporting Source Distribution for the Monthly Building Permit Survey

**Source:** U.S. Census Bureau; Building Permit Survey; Monthly Place Level Files

### 3.3 Survey Costs

Contrast the above information with Table 2 below showing reporting source distributions for the monthly BPS by the number of housing units authorized. Over this timeframe the percentage of permit-issuing places where missing data was obtained from another survey (i.e. SOC) rose from 3% (for places authorizing 0 housing units) to 60% (for places authorizing over 900 housing units). In these situations, missing data obtained from SOC adds burden and cost to the entire construction survey data collection process.

Reporting Source	Housing Units Authorized by the Monthly Building Permit Survey - January 2016 to January 2017							
	0 Units	> 0 Units	> 2 Units	> 15 Units	> 60 Units	> 150 Units	> 350 Units	> 900 Units
<i>Reports Received</i>	69%	61%	67%	69%	65%	62%	51%	38%
Paper Reporting	46%	33%	32%	29%	26%	26%	27%	28%
Electronic Reporting	4%	4%	3%	1%	2%	1%	1%	0%
Internet Reporting	20%	24%	32%	38%	37%	36%	24%	10%
<i>Reports not Received</i>	31%	39%	33%	31%	35%	38%	49%	63%
Data from SOC	3%	2%	3%	7%	15%	26%	42%	60%
Data from Imputation	26%	36%	30%	24%	20%	11%	7%	3%
Data not Received	2%	0%	0%	0%	0%	0%	0%	0%
<i>Sample Size</i>	53,817	21,177	22,574	8,697	2,320	902	356	80

**Table 2:** Reporting Source Distribution for the Monthly Building Permit Survey by Number of Housing Units Authorized

**Source:** U.S. Census Bureau; Building Permit Survey; Monthly Place Level Files

Taking the entirety of these issues discussed into consideration, the prospect of expending more resources to obtain residential data from other survey sources or increasingly burdened respondents takes on more importance.

## **4 New Data Sources**

### **4.1 Current Research**

In the past several years, permit-issuing places have begun making permit information available publicly. In fact, the BPS currently makes use of this information in limited situations. While the content, format, and accessibility of this information varies widely across places, it nonetheless makes sense to explore their viability as a new data source in more detail. As places move further away from traditional survey cooperation, this new survey reality may become more the rule than the exception.

This research serves just that purpose, to evaluate the use of publicly available building permit data to complement both data collection and analysis methods for the BPS and the SOC surveys. Lessons learned from this research can then be applied to additional data source formats moving forward.

### **4.2 Data Source Types**

The content, format, and accessibility of publicly available building permit information varies widely across places. Overall the various formats can be grouped into the following types: Application Programming Interfaces (APIs), building permit issuance reports, building permit issuance listings, and searchable building permit databases.

This research will initially center upon APIs with the hope that advantages and limitations uncovered will inform survey interactions with new data sources as they become available in the future.

## **5 Advantages & Limitations**

There are numerous advantages and limitations imposed by the incorporation of new data sources into current construction surveys. This research will discuss these advantages and limitations as they relate to issues regarding survey coverage, timeliness, completeness, classification, and validity. Together these issues may suggest additional research opportunities moving forward.

### **5.1 Coverage**

It is reasonable to assume that not every permit-issuing place will have publicly available building permit information. Online research for the largest permit-issuing places in SOC (based on total housing units authorized in 2015) has uncovered:

- 22 permit-issuing places having no public availability
- 32 permit-issuing places having a searchable building permit database
- 65 permit-issuing places having a building permit issuance report or listing
- 9 permit-issuing places having an API

These are by no means final numbers but lower bounds of expectation. Some permit-issuing places may only have general permitting information available publicly, while other places who do have publicly available building permit data may have failed to disclose this information.

It is also highly likely that certain types of public availability – like building permit APIs – would be more appropriate for permit-issuing places that issue large number of building permits. Permit-issuing places that issue limited numbers of building permits may not be willing to invest the necessary resources to incorporate this technology. This is interesting and suggests additional research opportunities moving forward.

## **5.2 Timeliness**

Frequency of data updates is an important limitation. In the case of monthly surveys, infrequent updates or potential missed updates due to human or computer error will introduce uncertainty and error into survey estimates. Investigation and documentation of update frequencies – daily, weekly, biweekly, monthly, etc. – is necessary.

API data sources appear to have the highest data update frequencies. Most APIs are routinely updated on a near-daily basis. Building permit issuance reports and listings have the most varied update frequencies. Availability of reports has been noted as weekly, monthly, quarterly, and annually.

## **5.3 Completeness**

The BPS and SOC are new, privately-owned, residential housing surveys that provide a range of data on residential housing topics. Topics covered by the surveys include number of buildings and housing units authorized by building permits, permit valuation, construction start and completion dates, construction characteristics (such as number of bedrooms, bathrooms, and square footage, etc.), and sales data (such as sales date, pricing, and financing).

Table 3 on the following page shows the types of residential data inputs needed for the BPS, and whether that data can be ascertained for new publicly available data sources, in this case permit-issuing places found to have an API. The availability of construction type (residential, commercial, etc.) and number of housing units are limited across all places. This will impact successful incorporation of these new data sources into BPS operations due to their significance in regards to survey estimates.

For example, Austin TX, Cary Town NC, Mesa AZ, and San Jose CA all appear to be excellent candidates for incorporation into BPS operations. The remaining 5 places all possess one or more limitations (reporting of construction classification or housing unit information) preventing incorporation:

Building Permit Survey	Publicly Available Data Sources - Application Programming Interfaces								
	Austin TX	Boston MA	Cary Town NC	Chicago IL	Las Vegas NV	Mesa AZ	Nashville TN	San Jose CA	Seattle WA
Residential Data Needed									
Type (residential, commercial, etc.)	Yes		Yes			Yes	Yes	Yes	Yes
Number of Buildings	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Housing Units	Yes	Occupancy	Yes			Dwelling		Dwelling	
Valuation	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes

**Table 3:** Availability of Residential Data for the Monthly Building Permit Survey via New Publicly Available Data Sources  
**Source:** U.S. Census Bureau; Building Permit Survey, How the Data are Collected & Publicly Available Data Sources

- Boston MA does not report classification (residential vs commercial).
- Nashville TN and Seattle WA do not report housing unit information.
- Chicago IL and Las Vegas NV do not report either type of information.

Table 4 below shows the types of residential data needed for the SOC and whether that data can be ascertained for new publicly available data sources, again for permit-issuing places found to have an API. The quantity and quality of overall residential data varies across places. Unlike BPS, the impact upon successful incorporation of these new data sources into SOC operations may be minimized.

Survey of Construction	Publicly Available Data Sources - Application Programming Interfaces								
	Austin TX	Boston MA	Cary Town NC	Chicago IL	Las Vegas NV	Mesa AZ	Nashville TN	San Jose CA	Seattle WA
Residential Data Needed									
Permit Information									
Permit Number	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Issue Date	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Housing Units	Yes	Occupancy	Yes			Dwelling		Dwelling	
Construction Information									
Construction Description	Yes	Yes	Yes	Yes			Yes	Yes	Yes
Address (Full, Partial, None)	Full	Full	Full	Partial			Full	Full	Partial
Construction Dates									
Start Date									
Certificate of Occupancy		Yes	Yes						
Completion Date	Yes		Yes						
Respondent Information	Applicant Contractor	Applicant	Contractor Owner	Contractor	Applicant		Contact	Applicant Contractor Owner	Applicant Contractor
Building Characteristics	Condo	Sq Foot	Sq Foot					Sq Foot	
> 25 Single-Family Items									
< 20 Multifamily Items	Floors								

**Table 4:** Availability of Residential Data for the Survey of Construction via New Publicly Available Data Sources  
**Source:** U.S. Census Bureau; Building Permit Survey, How the Data are Collected & Publicly Available Data Sources

For example, Boston MA and Cary Town NC both list certificate of occupancy information. This information is useful in the SOC as a means of estimating construction completion dates. In addition, information relating to condominium status, number of floors, and square footage are all characteristic information collected via the SOC. Incorporating this information may potentially increase data quality and lower survey respondent burden.

## **5.4 Classification**

As discussed above, permit-issuing places vary in reporting of housing unit and construction classification information. Housing unit information is a significant residential data item collected for the BPS and the SOC. When housing unit information is unavailable, use of construction classification information may suffice.

Three potential methods of assigning unavailable construction classification information (or when this information is incomplete or misclassified as per U.S. Census Bureau definitions) will be discussed – analyzing location outliers, permit valuation, and construction descriptions. These methods will suggest additional research opportunities moving forward.

### **5.4.1 Location Outlier Information to Assign Construction Classification**

The first potential method entails analyzing construction location information. Permit-issuing places enforce zoning laws that exclude (or limit) certain types of construction in certain geographic areas. Network analysis of known construction classifications using latitude/longitude coordinates (information which may be available on most new publicly available data sources) may serve this purpose.

### **5.4.2 Permit Valuation Information to Assign Construction Classification**

The second potential method entails analyzing construction valuation information. Valuation information is reported on most new publicly available data sources. Analysis of the valuation level provides an opportunity to classify construction as single-family or multifamily in nature, as it logically follows the valuation of a single-family house would in most cases be less than that of an apartment building, which would itself in most cases be less than that of a mixed use development.

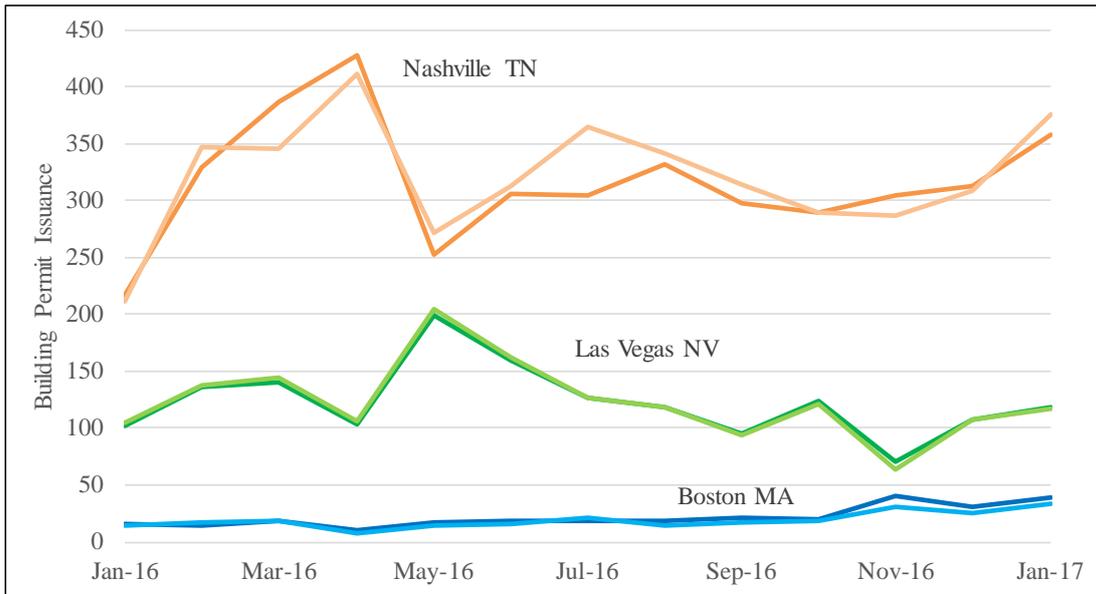
### **5.4.3 Work Description Information to Assign Construction Classification**

The final potential method entails analyzing the construction description. In these instances, developing a topic model via text analysis may allow us to assign a ‘modelled’ construction classification when this information is not available. For example, if the construction description references the keywords ‘apartment’ or ‘mixed-use’ a classification of multifamily residential would be appropriate. If the construction description references the keywords ‘townhouse’, ‘duplex’, or ‘rowhouse’ a classification of single-family residential would be appropriate.

## **5.5 Validity**

Prior to formally incorporating new publicly available data sources into BPS or SOC operations, data sources must be validated against current survey data to ensure all residential data associated with a permit-issuing place is accounted for. Against BPS data validation was conducted against the total number of buildings and total construction valuation reported between January 2016 and January 2017.

Figure 1 below shows the difference in building permit issuance levels between new publicly available data sources (in this case permit-issuing places of Boston MA, Las Vegas NV, and Nashville TN found to have an API) and BPS data from January 2016 to January 2017. As can be seen, issuance levels track closely for Boston and Las Vegas, and then begin to diverge for Nashville as monthly issuance levels increase.



**Figure 1:** Building Permit Issuance Levels Between new Publicly Available Data Sources - Application Programming Interfaces - and the Building Permit Survey

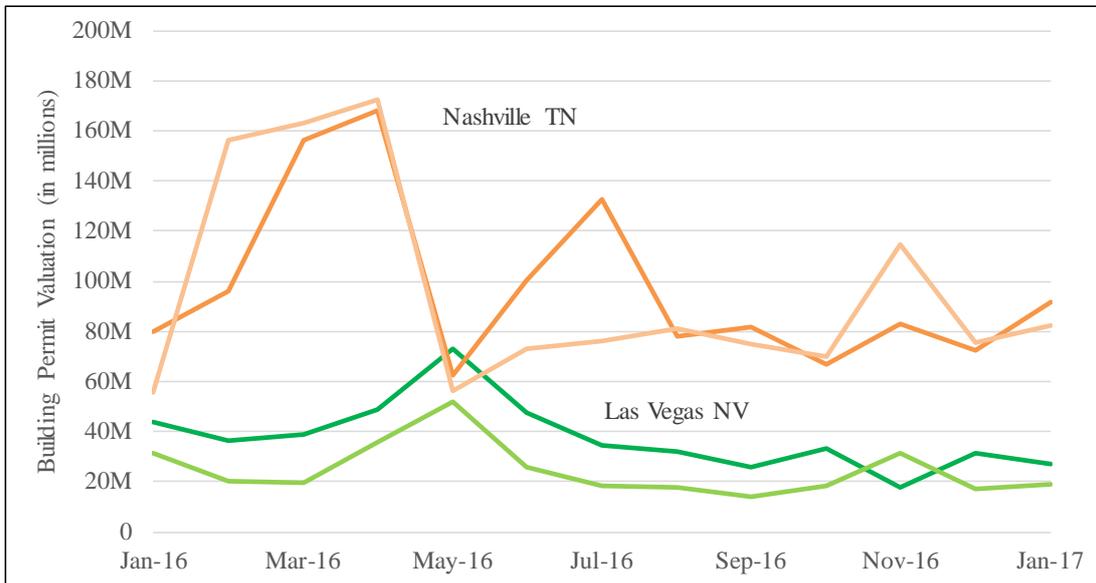
**Source:** U.S. Census Bureau; Building Permit Survey; Monthly Place Level Files & Publicly Available Data Sources

**Note:** For each place Building Permit Survey reporting source was 'Data from SOC' for all months

Figure 2 on the following page shows similar information for building permit valuation. Here valuation levels track differently for Las Vegas NV and Nashville TN. In Las Vegas valuations track similarly but with a consistent monthly difference, while in Nashville valuations are more varied across months. Both validations are interesting and suggest their use as complimentary data sources to other new construction surveys is promising.

## 6 Conclusions & Discussion

In summary, let us shift towards implications surrounding incorporating new data sources. The BPS and the SOC each collect similar information from permit-issuing place. Incorporating new publicly available data sources would serve to reduce this respondent burden. While contact may not be entirely eliminated, the reduction could serve to foster a better working relationship between permit-issuing places and the U.S. Census Bureau.



**Figure 2:** Building Permit Valuation Levels Between new Publicly Available Data Sources - Application Programming Interfaces - and the Building Permit Survey

**Source:** U.S. Census Bureau; Building Permit Survey; Monthly Place Level Files & Publicly Available Data Sources

**Note:** For each place Building Permit Survey reporting source was 'Data from SOC' for all months

Along similar lines, reduction in the frequency and/or complexity of survey requests to permit-issuing places could increase survey response rates. For example, assume it is determined new publicly available data sources can be substituted for the BPS, but the SOC must still collect data from respondents in permit-issuing places. One could suppose that a decrease in contact from the U.S. Census Bureau (no more monthly BPS!) may serve to increase, or at least hold steady, participation in the SOC.

What about data actually collected by surveys of new construction? As stated previously, data collection for establishment surveys entails specific respondent processing and reporting burdens that may negatively affect data quality. Incorporating new publicly available data sources may have the effect of reducing data collection errors and respondent burden (shorter SOC interviews!). In addition, these new data sources may allow for the analysis of new construction information not currently collected. This new information may serve to increase data quality or lend itself to the development toward new measurements of new construction.

Finally, data collection is a major component of survey costs. In instances where new publicly available data sources are incorporated, we may see a reduction in survey data collection costs. This would not be an absolute reduction as some data collection costs would transfer to other survey function. Exactly how much of an advantage these issues bring forth suggests additional research opportunities moving forward.

## References

Tony Hak, Diane K. Willimack & Amy Anderson (2003), “Response Process and Burden in Establishment Surveys”. Presented at the *American Statistical Association Joint Statistical Meetings*, San Francisco, CA: JSM – Section on Government Statistics, pp.1724-1730

Seymour Sudman, Diane K. Willimack, Elizabeth Nichols & Thomas Mesenbourg (2000), “Exploratory Research at the U.S. Census Bureau on Survey Response Process in Large Companies”. In *ICES II, Proceedings of the Second International Conference on Establishment Surveys*, Alexandria, VA: American Statistical Association, pp.327-339

Roger Tourangeau & Thomas Plewes (2013), “Nonresponse in Social Science Surveys: A Research Agenda”

U.S. Census Bureau. “Building Permits Place Level Files”. *Building Permit Survey*. Available at: <https://www2.census.gov/econ/bps/place> Accessed: June 2016.

U.S. Census Bureau. “How the Data are Collected”. *Building Permit Survey*. Available at: [http://www.census.gov/construction/bps/how\\_the\\_data\\_are\\_collected/](http://www.census.gov/construction/bps/how_the_data_are_collected/) Accessed: June 2016.

U.S. Census Bureau. “How the Data are Collected”. *Survey of Construction*. Available at: [http://www.census.gov/construction/nrc/how\\_the\\_data\\_are\\_collected/soc.html](http://www.census.gov/construction/nrc/how_the_data_are_collected/soc.html) Accessed: June 2016.

Application Programming Interface (API) for Boston, MA  
<https://data.cityofboston.gov/Permitting/Approved-Building-Permits/msk6-43c6>

Application Programming Interface (API) for Nashville, TN  
<https://data.nashville.gov/Licenses-Permits/Building-Permits-Issued/3h5w-q8b7>

Application Programming Interface (API) for Las Vegas, NV  
<https://opendata.lasvegasnevada.gov/Building-and-Safety/Building-Permits/bkdn-9ipc>