Mining TCAD and WCAD Public Record Data – ETL Guide to TCAD and WCAD

Michael McCarthy PE, Austin Energy, Austin TX USA

ABSTRACT

A multitude of Public Record Data (PRD) is readily available for SAS programmers interested in data mining government data. This paper presents a step-by-step tutorial on how to request Public Tax Record data, import the data into SAS®, and then perform ETL and data mining with BASE SAS®.

INTRODUCTION

The Texas Public Information Act (Texas General Code in Title 5, Subchapter A and Subtitle 552) requires that Texas government organizations respond to data inquiries from the general public. This means that you can submit a Public Information Request (PIR) to gain access to reports or data held by taxing organizations such as the Travis Central Appraisal District (aka TCAD www.traviscad.org) or Williamson Central Appraisal District (aka WCAD www.wcad.org). This data can be combined with your organization's existing customer data to help build contact lists, improve demographic profiles, or just provide a better understanding of physical properties of customer buildings. Here, ETL (Extract, Transform, and Load) is critical to accurately merging the public data set with your internal data set.

METHODOLOGY – HOW TO GET TCAD DATA:

Public Data is generally available via a PIR (Public Information Request). Requests should be sent to the Records Coordinator.

1.) Email the Records Manager (RecMgr@tcadcentral.org). Since this file is generally large, it's probably best to have them post to an SFTP server or a cloud drive (Google Drive, Drop Box, etc.)

> Hello TCAD Records Coordinator,
>
> <Your Company> would like to order the most recent copy of the TCAD property files as well as the most recent Shapefiles (STANDARD APPRAISAL EXPORT and GIS SHAPEFILES). Can you please send a PayPal invoice to <Contact Here> (I believe the fee is $80). <Contact Here> Email is <Contact Email Here>. Please use this link to upload the files to the SFTP file server (You will create a linked folder from your account to share SFTP LINK).
>
> - thanks, <Your Name Here>

2.) Here is a typical cost breakdown

Here is a breakdown of the costs:

Standard Appraisal Export will contain appraisal roll information for each parcel. The information is in ASCII comma delimited format and may be imported into the database program of your choice. A template and instructions for importing the data into Access is included on the CD. A file layout of all data fields and an improvement type list is also included. The cost of the Standard Appraisal Export is $55.

Itemized List of Charges:

Description	**Qty x Price**	**Total**
Labor minutes ($15/hour)	5 x $0.25	$1.25
Midsize minutes	35 x $1.50	$52.50
Client server minutes ($2.20/hour)	15 x $.0366	$0.55
Miscellaneous supplies	CD	$1.00
Total cost		**$55.00**

GIS Shape files are the outline of parcels. They contain only the PID # and the house number on some of the files. They do not contain the street name or owner name. You must have GIS or ESRI to use them. You also need to purchase the Export in order to have the data to match them to. The GIS data will match the appraisal roll on a one to one basis. However, the GIS data is typically not as up-to-date as information extracted from our appraisal roll. We use the Prop ID field to make all joins. This is the key link between property ownership and GIS information. The export CD should include instructions for importing the text files into an access database. Once the files have been imported you can join the owner name and address data. The cost of the Shape file CD is $25.

Itemized List of Charges:

Description	**Qty x Price**	**Total**
Labor minutes ($15/hour)	5 x $0.25	$1.25
Midsize minutes	15 x $1.50	$22.50
Client server minutes ($2.20/hour)	5 x $.0366	$0.18

3.) The tables will be in an Access database format. Unpacking instructions are included.
   a. The tables used in the database are: IMP_DET, IMP_INFO, LAND_DET, PROP, STATE_CD
   b. Also download the most recent 'GIS Data'. Download to a network share or your local drive or order the CD

HOW TO GET WCAD DATA

4.) Ownership information can only be obtained via a PIR. There is no cost at this time. Send the following Email as a PIR to WCAD (PIR@wcad.org)

Hello WCAD,

My name is <Your Name Here>. I work for the <YOUR COMPAY NAME HERE>. I'm requesting a copy of the all unrestricted fields from the "Owner" table within the 2018 Access Database file. I would please like this in one of the following standard electronic formats: dbf, mdb, csv, txt, xlsx. A csv file is fine if this is easiest. I'm requesting this information so that I can update our databases with new property ownership information. Please let me know if you have any questions regarding this request

- Thanks, <Your Name Here>

5.) The remaining WCAD tables can be downloaded directly from the WCAD website:
   a. Go to http://www.wcad.org/ and scroll down to the 'Data Downloads' section. Click 'Learn More' and 'accept disclaimer'
   b. You can choose 'Certified Data' and download the '<Year> Access Export' (or possibly '<Year> Property Data Export')
   c. If you select 'Non-Certified' data, you can get an Access version of next year's data
   d. Download and UNZIP the file.
   e. The tables used in the database are: Property, Owner, Sales, ImpSeg – Real Property, Improvement – Real Property, Final Values
   f. Also download the most recent 'GIS Data':
      i. Select GIS Data 'Parcels'
      ii. Unzip the Parcel file. This is the Shapefile database.
   g. Download to a network share or your local drive.

NOW UNZIP BOTH FILES AND UNPACK THE DATABASES

6.) TCAD

a. Once the fee has been paid, the TCAD coordinator will transfer the .zip files. Follow the "TCAD_Roll_Extract_Documentation" to rebuild the Access database.
b. Follow the unpacking instructions for the TCAD files.
c. The shapes files contain a map called "Parcel_Poly.shp". This map is in the Texas Planar (NAD 1983 State Plane Texas Central FIPS 4203 Feet) coordinate system (ESRI:102739). Details below:

Project your lat long onto NAD 1983 State Plane Texas Central FIPS 4203 Feet Parameters: +proj=lcc +lat_1=30.11666666666667 +lat_2=31.88333333333333 +lat_0=29.66666666666667 +lon_0=-100.3333333333333 +x_0=700000 +y_0=3000000 +ellps=GRS80 +datum=NAD83 +to_meter=0.3048006096012192 no_defs

7.) WCAD

a. Download the WCAD Access database: "<Year> Appraisal Roll Report"
b. Perform a compact and repair.
c. Import the "OWNER" Table into the WCAD database
d. The shapes files contain a map called "Parcel_Poly.shp". This map is in the Texas Planar (NAD 1983 State Plane Texas Central FIPS 4203 Feet) coordinate system (ESRI:102739). Details below:

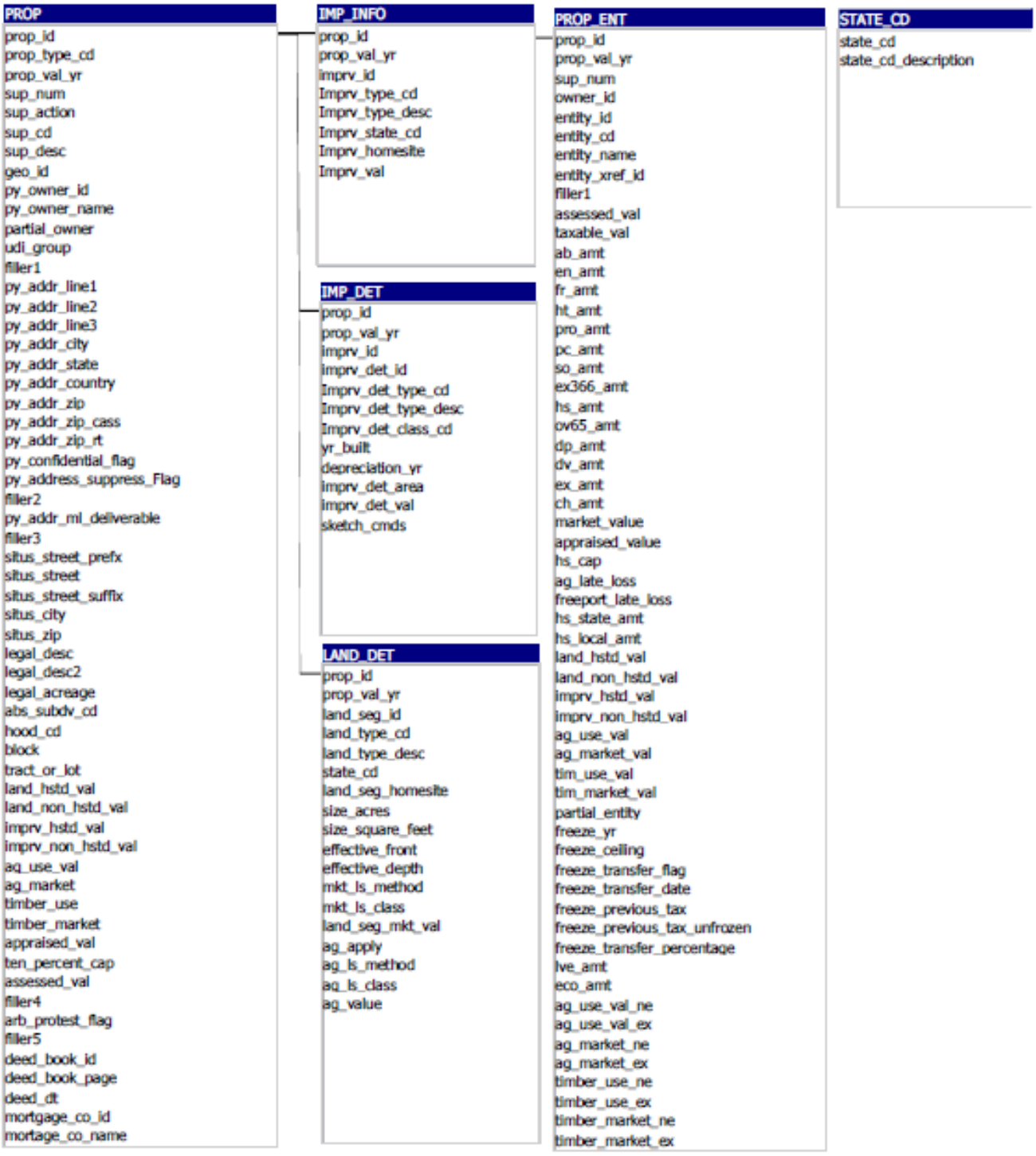
Project your lat long onto NAD 1983 State Plane Texas Central FIPS 4203 Feet Parameters: +proj=lcc +lat_1=30.11666666666667 +lat_2=31.88333333333333 +lat_0=29.66666666666667 +lon_0=-100.3333333333333 +x_0=700000 +y_0=3000000 +ellps=GRS80 +datum=NAD83 +to_meter=0.3048006096012192 no_defs

It's important to be aware of the planar coordinate system of your shapefiles so that you can accurately overly other SAS® coordinate data with these. The example below shows how to intersect GPS data with your shapefile data using the sashelp.proj4def data set:

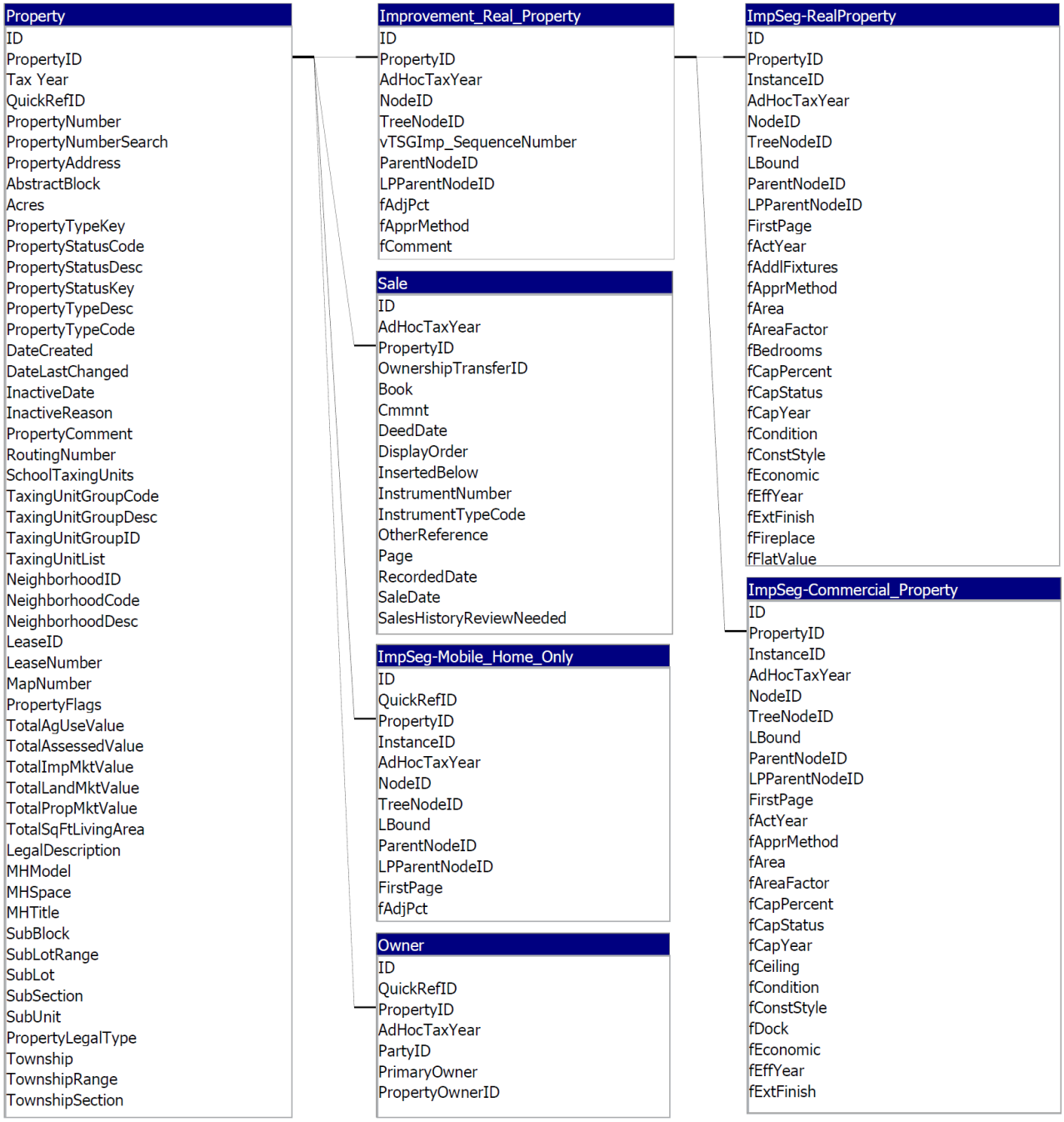
```
PROC MAPIMPORT OUT=WCAD DATAFILE="YOUR FILE PATH HERE\Parcel_poly.shp";
run;

/*DEFINE LATTITUDE, LONGITUDE, AND POINT ID*/
data have (KEEP = lat long PointID);
      set <YOURPOINT DATA SET HERE>;
      run;

/*PROJECT YOUR DATA ONTO THE TEXAS PLANAR COORDINATE SYSTEM*/
proc gproject latlon
project=proj4 to="ESRI:102739"
data=have out=want;
id PointID;
run;

/*PERFORM A GEOSPACIAL OVERLAY - THIS STEP CAN TAKE SOME TIME ☹*/
proc ginside data=WANT map=WCAD out=FOUND INSIDEONLY;
  id OBJECTID PROP_ID;
run;
```

RELATIONSHIPS BETWEEN FIELDS AND TCAD DATA TABLES

In the TCAD database you've just unpacked, the "prop_id" field serves as the "primary key". This key makes it possible to relate each table to the other tables. While there are many tables available in the TCAD public record database, I've found the tables listed below to be the most useful:



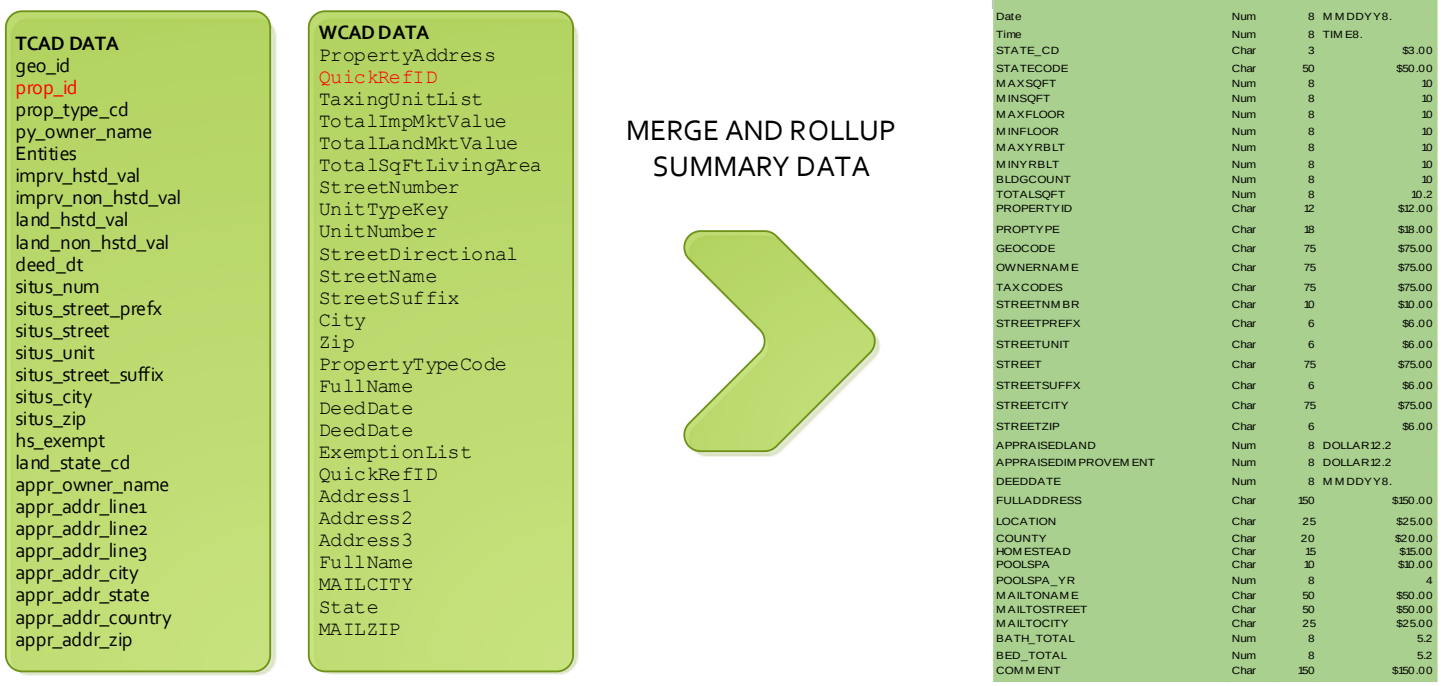
Similarly, in the WCAD database you've just unpacked, the "PropertyID" field serves as the "primary key". While there are many tables available in the WCAD public record database, I've found the tables listed below to be the most useful:



Interpreting and translating database fields can sometimes be more art than data science. Be sure to have a solid understanding of what the field values represent.

DATA FLOW DIAGRAM

Once you've imported the TCAD or WCAD data into SAS, you can then think about a data flow diagram outlining how you want to compose the data so that it will make sense within your organization. An ETL Data flow shown below:



USING SAS CODE TO TRANSFORM YOUR DATA

Now that you have the data in a Microsoft ACCESS® format, you can create a BASE SAS® library to extract the fields you need for your project:

```
/*INPUT DATA LIBRARIES – USING SAS/ACCESS Interface to PC Files*/
libname TCADIN 'C:\LOCATION OF DATABASE\TCAD.accdb';/*TCAD IN*/
ibname WCADIN 'C:\LOCATION OF DATABSE\WCAD.accdb'; /*WCAD IN*/
/**OUTPUT DATA LIBRARIES**/
/*TCAD*/libname TCADOUT "C:\OUTPUT DIRECTORY\TCAD";
/*WCAD*/libname WCADOUT "C:\OUTPUT DIRECORY\WCAD";
```

You can now begin combining the fields you need for your project. Use the "PropertyID" or "PROP_ID" to combine fields from different tables. Here is an example SQL macro which will do this…

```
/*************************************************************/
/*                        JOINTBL Macro                      */
/* MERGE TWO SOURCE TABLES AND OUTPUT A SAS SOURCE TABLE     */
/*************************************************************/
%macro JOINTBL(INDIR,INTABLE1,INTABLE2,OUTDIR,OUTTABLE,FIELD1,FIELD2);
proc sql dquote=ansi;
      create table &OUTDIR..&OUTTABLE as
      select distinct a.*,b.*
      from &INDIR..&INTABLE1 as a left join &INDIR..&INTABLE2 as b on a.&Field1=b.&Field2
      quit;
%mend JOINTBL;
Call EXECUTE ('%JOINTBL(WCADIN,"Improvement_Real_Property","ImpSeg-
RealProperty",WCAD,JOINCODES,PropertyID,PropertyID)');
```

It's helpful to create descriptive formats using "PROC FORMAT" in SAS to keep you on track while summarizing and defining your final data output. This can make the ETL steps less confusing…

```
proc format; /*MAPPING THE INPUT AND OUTPUT FIELDS*/
        value $FieldLabel
        /*TCAD*/ geo_id                 = GEOCODE
                  py_owner_name         = OWNERNAME
                  entities              = TAXCODES
                 appraised_val          = APPRAISED
                 deed_dt                = DEEDDATE
                 situs_num              = STREETNMBR
                 situs_street_prefx     = STREETPREFX
                 situs_street           = STREET
                 situs_unit             = STREETUNIT
                 situs_street_suffix    = STREETSUFFX
                 situs_city             = STREETCITY
                 situs_zip              = STREETZIP
         /*WCAD*/  TaxingUnitList       = TAXCODES
                 TotalAssessedValue     = APPRAISED
                 TotalSqFtLivingArea    = TOTALSQFT
                 StreetNumber           = STREETNMBR
                 StreetDirectional      = STREETPREFX
                 StreetName             = STREET
                 UnitNumber             = STREETUNIT
                 StreetSuffix           = STREETSUFFX
                 City                   = STREETCITY
                 Zip                    = STREETZIP; run;
```

```
proc format;/*HOW TO COUNT FLOORS*/
        value $FloorCount
        '1st Floor'          = 'COUNT' /*TCAD FIELD*/
        '2nd Floor'          = 'COUNT' /*TCAD FIELD*/
        '3rd Floor'          = 'COUNT' /*TCAD FIELD*/
        '4th Floor'          = 'COUNT' /*TCAD FIELD*/
        '5th Floor'          = 'COUNT' /*TCAD FIELD*/
        '6th Floor'          = 'COUNT' /*TCAD FIELD*/
        'Lobby'              = 'COUNT' /*TCAD FIELD*/
        'Additional Floor'   = 'COUNT' /*TCAD FIELD*/
        'MA'                 = 'COUNT' /*WCAD FIELD*/
        'MA2'                = 'COUNT' /*WCAD FIELD*/
        'MA3'                = 'COUNT' /*WCAD FIELD*/; run;

proc format;/*HOW TO FIND POOLS*/
        value $WaterCode
/*POOLS*/'601'   =   "POOL COMM'L"
         '609'   =   'POOL FV'
         '604'   =   'POOL RES CONC'
         '603'   =   'POOL RES FIBERGL'
/*SPAS*/ '349'   =   'SPA FV'
         '448'   =   'SPA FIBERGLASS'
         '447'   =   'SPA CONCRETE'
         '449'   =   'SPA'
         '1'     =   'SCP'; run;
```

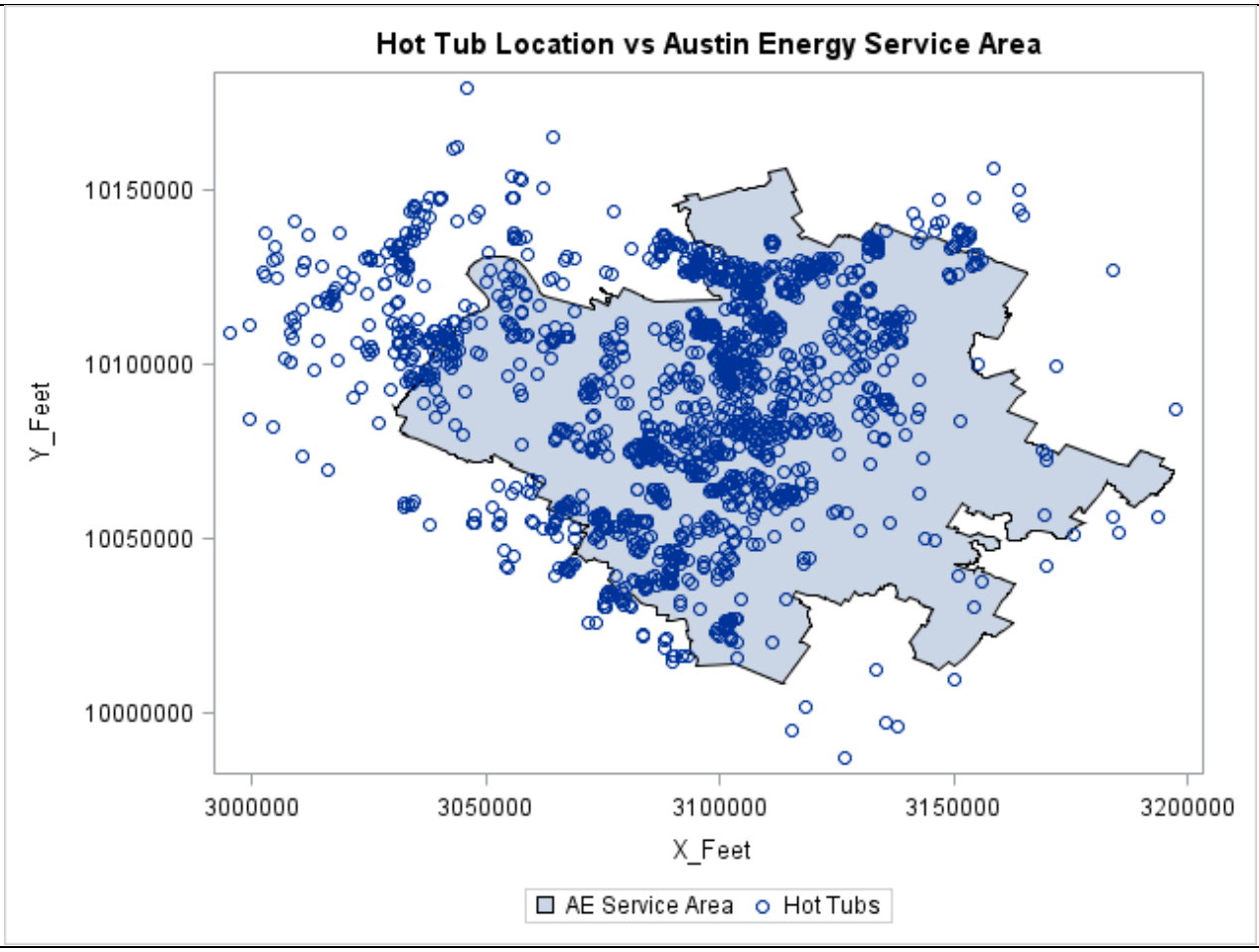
```
proc format;/*StateCode*/
        value $StateCode
/*TCAD FIELDS*/
'A1'    =   'SINGLE FAMILY RESIDENCE'
'A2'    =   'SINGLE FAMILY RESIDENCE MH'
'A3'    =   'SINGLE FAMILY RESIDENCE DETAILS'
'A4'    =   'CONDOS'
'A5'    =   'CONDOS DETAILS'
'A9'    =   'HS COMMERCIAL HIGHEST & BEST USE'
'B1'    =   'MULTIFAMILY'
'B2'    =   'DUPLEX'
'B3'    =   'TRI-PLEX'
'B4'    =   'FOUR-PLEX'
'C1'    =   'VACANT LOT'
'C2'    =   'VACANT LAND/MISC DETAILS'
'D1'    =   'ACREAGE (AG) 1-D-1'
'D2'    =   'ACREAGE (NON-AG)'
'D3'    =   'AG 1-D'
'E1'    =   'FARM AND RANCH IMPR'
'E2'    =   'FARM AND RANCH IMPR MH'
'E3'    =   'FARM AND RANCH IMPR MISC'
'F1'    =   'COMMERCIAL IMPROVED'
'F2'    =   'INDUSTRIAL MAJOR MANUFACTURING IMPROVED'
'F3'    =   'COMMERCIAL DETAILS'
'F4'    =   'COMMERCIAL CONDO'
'F5'    =   'COMMERCIAL RES CONVERSION'
'G1'    =   'MINERAL'
'G2'    =   'Mineral'
```

```
'G3'    =   'Mineral'
'J1'    =   'UTILITY (WATER)'
'J2'    =   'UTILITY (GAS)'
'J3'    =   'UTILITY (ELECTRIC)'
'J4'    =   'UTILITY (TELEPHONE)'
'J5'    =   'UTILITY (RAILROADS)'
'J6'    =   'UTILITY (PIPELINES)'
'J7'    =   'UTILITY (CABLE)'
'J8'    =   'UTILITY (OTHER)'
'J9'    =   'UTILITY (NOT CODED)'
'L1'    =   'COMMERCIAL PP'
'L2'    =   'INDUSTRIAL MAJOR MANUFACTURING PP'
'L3'    =   'Mineral'
'M1'    =   'TANGIBLE PERSONAL PROP MH'
'M2'    =   'NON-INCOME PRODUCING TANGIBLE PP'
'N'     =   'Mineral'
'N1'    =   'INTANGIBLE PP'
'N2'    =   'RR ROLLING STOCK'
'O1'    =   'RESIDENTIAL INVENTORY'
'S1'    =   'SPECIAL INVENTORY'
'X'     =   'TOTALLY EXEMPT PROPERTY'
'Y'     =   'Mineral'; run;
```

GETTING INSIGHTS FROM YOUR DATA

The example below shows how you can use public data to better understand your customers. The BASE SAS® code below will overlay properties with hot tubs on the Austin Energy Service Territory. Going further, you could use the "PROC GINSIDE" procedure to get a good count of all customers with hot tubs.

```
/*************************************************************************/
/* SCSUGS 2018 - DEMO MAP OF HOT TUBS IN AUSTIN ENERGY SERVICE TERRITORY
/*************************************************************************/
PROC MAPIMPORT OUT=AE_Service DATAFILE="\\LOCATION OF SHAPEFILE\Austin_Energy_Service_Area.shp";
run;
PROC MAPIMPORT OUT=TCAD DATAFILE="\\LOCATION OF SHAPEFILE\Parcel_poly.shp";run;
libname WTCAD '\\LOCATION OF YOUR TCAD AND WCAD DATA PROPERTY DATA';
Data SPA (KEEP=PROPERTYID POOLSPA POOLSPA_YR);
        set WTCAD.TCADWCAD; /*List of prop_ids in Travis County with only spas: 1,682 found*/
        if POOLSPA = 'SPA' then output; run;
proc sql;
        create table SPA as
        select distinct
        a.*,b.X,b.Y,b.OBJECTID,PROP_ID
        from SPA as a inner join TCAD as b on
             a.PROPERTYID = PUT(b.prop_id, best6.);/*prop_id is an integer while PROPERTYID is a character ☹*/
        quit;
proc sort data=SPA NODUPKEY; by OBJECTID; run;/*CREATE X,Y LIST OF POINTS - KEEP ONLY ONE POINT FROM EACH PLAT*/
/*DEFINE X,Y FOR SCATTERPLOT*/
data SPA (DROP = X Y);
        set SPA;
        X_Feet = x; Y_Feet=y;run;
/*MERGE THE DATA SETS*/
data combined; set SPA AE_Service; run;
/*Plot Hot Tubs in Travis County over Austin Energy Service Territory*/
proc sgplot data=combined ;
  TITLE "Hot Tub Location vs Austin Energy Service Area";
  polygon x=x y=y ID=SYSTEM ID / fill outline LEGENDLABEL='AE Service Area';
  scatter x=X_Feet y=Y_Feet / LEGENDLABEL='Hot Tubs';run;
```



CONCLUSION

The instructions presented in this paper should get you started with TCAD and WCAD public record data. This type of data is very useful when establishing legal property ownership, taxable property attributes, and owner mailing information. You can add this data into existing databases, and relate it to your existing data at your organization using geocoding or GIS tools such as ESRI ARCGIS®. Happy Coding ☺

ACKNOWLEDGEMENTS

Special thanks to John Trowbridge, PE (DABI Engineering Supervisor, Austin Energy), Liz Jambor (DABI Manager, Austin Energy), Victor Carr (Division Manager of Information Systems, Austin Energy) and Debbie Kimberly (VP Customer Energy Solutions, Austin Energy) for their continued support of our SAS Data Analytics Server.

REFERENCES

Texas Public Information Act
https://en.wikipedia.org/wiki/Texas_Public_Information_Act

Paper BT005 Maps Made Easy Using SAS®
Mike Zdeb, University of Albany School of Public Health
https://www.lexjansen.com/nesug/nesug03/bt/bt005.pdf

Maps Made Easy Using SAS®
Mike Zdeb, Art Carpenter, July 2002
https://www.sas.com/store/books/categories/getting-started/maps-made-easy-using-sas-/prodBK_57495_en.html

CONTACT INFORMATION

I value your comments. You can contact the author at:

Michael McCarthy, PE
Austin Energy, Customer Energy Solutions, Austin TX USA
michael.mccarthy@austinenergy.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.