Indexed Data Standardization Guide

Developing a Data System to Address Foreclosures in Massachusetts

Team Members

John McCarthy, John Riley, Natalie Dionne, Sophia Noel, And Theodore Campbell

Project Advisors

Dr. Scott Jiusto

Dr. Gbetonmasse Somasse

Project Liaison

Grace Ross

Massachusetts Alliance Against Predatory Lending (MAAPL)

and Worcester Anti-Foreclosure Team (WAFT)

December 10, 2020

WPI Cape Town Project Center

Please view our website for more information on our project:
https://wp.wpi.edu/southafrica/projects/2020-projects/data/

Project Summary:

This project assisted the Mass Alliance Against Predatory Lending (MAAPL) in addressing the ongoing home foreclosure crisis by gathering data into an organized data system that could be used to analyze trends in foreclosures in the state. To this end, we devised a plan to accumulate data for our sponsor which resulted in a framework of a data system to support future analyses. The data system includes instructional guides for gathering the information, a functioning database, and data that has been gathered from public sources. We also developed a guide for future development of the system. The project also served to develop an initial relationship between WPI and MAAPL.

Document Description

The Registries of Deeds throughout Massachusetts index and store data in a variety of ways and using a variety of database management systems. Because of this, the format of the indexed data often differs depending on the Registry. This document details some tips to use when collecting indexed data and modifying it into a standardized format from each of the seven Registries that the team has received data files from. These techniques can also be used to modify the data reformatting process based on data collected from other registries in the future. Each data file must be edited into the same standardized format, so that the CSV files are easily added to the database and the data go to the correct tables within the database. The format of the data input into the fields must also be standardized so that data can be entered into and queried from the database.

Data Field Formatting Conventions

The suggested CSV data file format is listed below, along with an example of how these fields should be laid out in a CSV file and was determined by leveraging the useful data fields received from a variety of counties in Massachusetts. The goal is to maximize all data that can be collected and entered into the database. Because some registries collect and index data that other registries do not, there will be some fields that are completely blank for certain registries. The database can interact with blank fields, so these blanks present no problems for the data input and storage system.

Data Field	Formatting Convention
Document ID	A string- combination of numbers, letters, and spaces (calculated)
Foreclosure ID	A string- combination of numbers, letters, and spaces (calculated)
Document Number	String of numbers, no spaces
Document Type	"FORECLOSURE DEED", "CERTIFICATE OF ENTRY", and/or "POWER OF ATTORNEY"
Town	Full or partial town name
Book	String of numbers, no spaces
Page	String of numbers, no spaces
Recorded Date	mm/dd/yyyy
Recorded Time	hh:mm (24 hour format)
Consideration	Price with money sign, ex: $100,000
Document Date	mm/dd/yyyy
Street Number	Number
Street Name	Regular street name with spaces and conventional abbreviations, ex: SCHOOLHOUSE RD
Land Type	Recorded (REC) or Registered (REG)
Additional Info	Property/Document description (there usually isn't both), text description or number code
Grantor 1-12	Party Name: Last Name First name Middle initial (no commas) OR bank name
Grantee 1-12	Party Name: Last Name First name Middle initial (no commas) OR bank name

Document ID	Foreclosure ID	Document Num	Document Type	Town	Book	Page	Recorded Da	Recorded	Consideration	Document Street Num	Street Name	Land Type	Additional
BOSTON5256387FORECLOSURE DEED	4164675CHESTNUT ST UNIT 13BOSTON	1208	FORECLOSURE DEED	BOSTON	52563	87	1/7/2014		402000.00		75	CHESTNUT S REC	0
BOSTON5256386CERTIFICATE OF ENTRY	4164675CHESTNUT ST UNIT 13BOSTON	1207	CERTIFICATE OF ENTRY	BOSTON	52563	86	1/7/2014		0.00		75	CHESTNUT S REC	0
REVERE5256836FORECLOSURE DEED	41648632REVERE BEACH BLVDREVERE	1562	FORECLOSURE DEED	REVERE	52568	36	1/9/2014		456000.00		632	REVERE BEA REC	0
BOSTON5256896FORECLOSURE DEED	416484DOVE STBOSTON	1574	FORECLOSURE DEED	BOSTON	52568	96	1/9/2014		184500.00		4	DOVE ST REC	DORCHEST
REVERE5256835CERTIFICATE OF ENTRY	41648632REVERE BEACH BLVDREVERE	1561	CERTIFICATE OF ENTRY	REVERE	52568	35	1/9/2014		0.00		632	REVERE BEA REC	0
BOSTON5256895CERTIFICATE OF ENTRY	416484DOVE STBOSTON	1573	CERTIFICATE OF ENTRY	BOSTON	52568	95	1/9/2014		0.00		4	DOVE ST REC	DORCHEST
CHELSEA52582274FORECLOSURE DEED	4165456-58CRESCENT AVECHELSEA	2923	FORECLOSURE DEED	CHELSEA	52582	274	1/15/2014		261071.21		56-58	CRESCENT A REC	0
CHELSEA52582272CERTIFICATE OF ENTRY	41654CHELSEA	2922	CERTIFICATE OF ENTRY	CHELSEA	52582	272	1/15/2014		0.00	0	0	REC	0
BOSTON52595171FORECLOSURE DEED	41656278RPRINCETON STBOSTON	3922	FORECLOSURE DEED	BOSTON	52595	171	1/17/2014		100000.00		278R	PRINCETON REC	EAST BOST
BOSTON52595170CERTIFICATE OF ENTRY	41656278RPRINCETON STBOSTON	3921	CERTIFICATE OF ENTRY	BOSTON	52595	170	1/17/2014		0.00		278R	PRINCETON REC	EAST BOST
REVERE52597185FORECLOSURE DEED	41660395LEE BURBANK HGWY UNIT 12REVER	4078	FORECLOSURE DEED	REVERE	52597	185	1/21/2014		313842.57		395	LEE BURBAN REC	0
REVERE52597184CERTIFICATE OF ENTRY	41660395LEE BURBANK HGWY UNIT 12REVER	4077	CERTIFICATE OF ENTRY	REVERE	52597	184	1/21/2014		0.00		395	LEE BURBAN REC	0
BOSTON5262198FORECLOSURE DEED	4166610CHENEY STBOSTON	6858	FORECLOSURE DEED	BOSTON	52621	98	1/27/2014		250000.00		10	CHENEY ST REC	DORCHEST
BOSTON5262717180FORECLOSURE DEED	4166611-13WEBSTER ST UNIT 3BOSTON	6341	FORECLOSURE DEED	BOSTON	52617	180	1/27/2014		360926.16		11-13	WEBSTER ST REC	HYDE PAR
BOSTON5262197CERTIFICATE OF ENTRY	4166610CHENEY STBOSTON	6857	CERTIFICATE OF ENTRY	BOSTON	52621	97	1/27/2014		0.00		10	CHENEY ST REC	DORCHEST
BOSTON52617178CERTIFICATE OF ENTRY	4166611-13WEBSTER ST UNIT 3BOSTON	6340	CERTIFICATE OF ENTRY	BOSTON	52617	178	1/27/2014		0.00		11-13	WEBSTER ST REC	HYDE PAR
REVERE52625157FORECLOSURE DEED	4166776JOHNNY RDREVERE	7380	FORECLOSURE DEED	REVERE	52625	157	1/28/2014		223000.00		76	JOHNNY RD REC	0
WINTHROP52622239FORECLOSURE DEED	4166719-21CROSS STWINTHROP	7065	FORECLOSURE DEED	WINTHRO	52622	239	1/28/2014		419000.00		19-21	CROSS ST REC	0

Above: an example of a standardized indexed data sheet from the Suffolk County Registry of Deeds

Unique Data Formats

The different Registries of Deeds in Massachusetts store and export data in different formats. For example, some registries report the names of both party types that are involved in the foreclosure (grantor and grantee) in the same row. This document preserves the relationship between pairs of parties and condenses the data file by including more than one name in a row.

For example, as seen below, the Dukes County Registry of Deeds provides indexed data that follows this format, where Column B indicates the name of the grantor for the foreclosure document described in that line, and Column C indicates the name of the grantee for that foreclosure document.

A	B	C	D	E	F	G	H	I
Document Number	Grantor	Grantee	Type	Town	Blank	Book	Page	Recorded Date	Reco
220	TITAN CAPITAL INC	TITAN CAPITAL INC	FORECLOSURE DE	EDGARTOWN		1339	513	1/13/2014
220	LLMJ INVESTORS LLC	TITAN CAPITAL INC	FORECLOSURE DE	EDGARTOWN		1339	513	1/13/2014
276	MARTHAS VINEYARD SAVINGS BANK	JACOBS HERBERT	FORECLOSURE DE	EDGARTOWN		1339	850	1/15/2014
276	DUKES COUNTY SAVINGS BANK	JACOBS HERBERT	FORECLOSURE DE	EDGARTOWN		1339	850	1/15/2014
276	STARBUCKS RIDGE REALTY TRUST	JACOBS HERBERT	FORECLOSURE DE	EDGARTOWN		1339	850	1/15/2014
276	CIRCLE EE REALTY TRUST	JACOBS HERBERT	FORECLOSURE DE	EDGARTOWN		1339	850	1/15/2014
276	HALL BENJAMIN L JR TRS	JACOBS HERBERT	FORECLOSURE DE	EDGARTOWN		1339	850	1/15/2014
452	BANK OF AMERICA NA	SAUCIER LANA S	FORECLOSURE DE	OAK BLUFFS		1340	471	1/24/2014
452	SHELLEY DENNIS L	SAUCIER LANA S	FORECLOSURE DE	OAK BLUFFS		1340	471	1/24/2014
452	SHELLEY CATHERINE	SAUCIER LANA S	FORECLOSURE DE	OAK BLUFFS		1340	471	1/24/2014
1059	NATIONSTAR MORTGAGE LLC	NATIONSTAR MORTGAGE LLC	FORECLOSURE DE	OAK BLUFFS		1342	603	2/25/2014
1059	REITZAS MORGAN	NATIONSTAR MORTGAGE LLC	FORECLOSURE DE	OAK BLUFFS		1342	603	2/25/2014

Above: a section of Dukes County original recorded land data

On the other hand, other Registries of Deeds report only one party name per row and specify which party type the name refers to. This creates more lines in the data file per document but simplifies the columns. See below for an example from the Northern Middlesex County Registry of Deeds, where column D indicates the party type (grantor= D or grantee= I) and the name columns (columns F and G) indicate the last and first names of the party.

	A	B	C	D	E	F	G	H	I	J	K	L
3	8/5/2015	POSSN		I	283320	CITIMORTGAGE INC				BILLERICA	194	77
4	8/5/2015	POSSN		D	283320	LUCAS	ADAM			BILLERICA	194	77
5	8/5/2015	POSSN		D	283320	LUCAS	CHRISTOPHER			BILLERICA	194	77
6	8/28/2015	POSSN		D	283724	NEAL	STEPHEN			BILLERICA	192	393
7	8/28/2015	POSSN		D	283724	NEAL	DEBORAH A			BILLERICA	192	393
8	8/28/2015	POSSN		I	283724	GREEN TREE SERVICING LLC				BILLERICA	192	393
9	3/4/2016	POSSN		D	285969	TAYLOR	GARY			BILLERICA	172	81
10	3/4/2016	POSSN		I	285969	DEUTSCHE BANK NATIONAL TRUST CO TR				BILLERICA	172	81
11	3/4/2016	POSSN		D	285969	SONNENB LINDA				BILLERICA	172	81

Above: Section of original Middlesex County indexed data sheet

In our standardization form, we worked to create format similar to the Dukes County Registry of Deeds, so that the information related to one document is condensed into one line of data in the reformatted file. This new format creates a consolidated data file that can be used to easily input data into the database and can be easily read by the human eye as well. We created a system to automatically reformat files in the format above (like Middlesex North), because that is the most common format, but the system cannot yet deal with formats like Dukes County above. That is why data from some Registries can be standardized using this Excel file, but other Registries will require a different or revised system.

Indexed Data Standardization Workbook Overview

The Indexed Data Standardizing Workbook is currently optimized for use with the data from the following Registries of Deeds:

- Berkshire Middle Registry of Deeds
- Bristol North Registry of Deeds
- Middlesex North Registry of Deeds
- Suffolk Registry of Deeds

Data from the following three Registries of Deeds are not yet compatible with this standardization workbook, due to the complexity of the data files we received from them:

- Southern Essex Registry of Deeds
  - Town names are abbreviated, making it difficult to input complete addresses
  - No party names
- Dukes County Registry of Deeds
  - Non-standard format
- Nantucket County Registry of Deeds
  - Non-standard format

In addition, the workbook is only compatible with indexed data for Foreclosure Deeds, Certificates of Entry, and Affidavits. It is not currently compatible with Power of Attorney indexed data. Powers of Attorney often include up to 20 Grantor and Grantee names on it, which is too many for the current workbook to store and reformat without risking losing data. Changing the programming on the Excel workbook would be helpful to optimize the storage of many grantor and grantee names and therefore make the workbook compatible with documents that have many grantor and grantee names, like the Power of Attorney documents.

Using the Indexed Data Standardizing Workbook

For the document types from the four Registries that are compatible, we recommend following the procedure below to best utilize the Indexed Data Standardizing Workbook:

1. Input the data into the corresponding sheet in the Standardizing Workbook:
   - Select the entire top row of the data (not including the headings) by pressing and dragging the curser across the row to highlight it, pressing the "control", "shift", down arrow keys at the same time to select the entirety of the data set.
   - Copy the data by pressing "control" and "c" at the same time.
   - Pull up the sheet in the Standardizing Workbook that is titled "[Registry Name] calc".
   - Paste the data into the corresponding sheet by selecting the top-left most cell below the data column names (see below- select the highlighted cell) and selecting "control"-"v" at the same time to enter the original data into the workbook
   - The data should be automatically formatted as an Excel table. You can tell if it is a table by the small down arrows in each of the document headings. Also, a tab titled "Table Design" will be shown in the upper green ribbon in the Excel sheet.
     - If the data is not automatically in a table, select the entirety of the data and press "control" and "T" at the same time to make the data into a table. In the dialog box select the "my table has headers" box and press "OK". The right side of the sheet should then automatically populate itself into a standardized version of the original data.

Original Middlesex North Data:

date recorded	document type	consideration	party type	document number	last name	first name	street num	street name	town	book	page

2. Remove Power of Attorney or other data if needed:
   - To remove Power of Attorney data or other data that is not compatible with the workbook, if not already completed, sort the original data table by document type.
   - To do this, press the small downward arrow to the right of the column that contains the document types (usually labelled "document type", "type" "descr" or any other similar name) and press the "Sort A to Z" option. This will rearrange the data so that you can copy the data you do not want into a separate sheet for later use. Use the copy and paste method described earlier and delete the cells from the "calc" sheet once it has been entered into the extra sheet.

3. Ensure that the data has been correctly populated and formatted:
   - Scroll to the right, where the data should be repopulated into the corresponding reformatted table. There should be names in some of the grantor and grantee columns.

- If there are any issues, check the logic of the Excel code to ensure it is rearranging the data correctly.

4. <u>Move the reformatted data:</u>
   - Using a similar copy and paste method as described in step 1, copy the data in the reformatted sheet to the corresponding "final" data sheet. However, when pasting, paste only the data values and not the equations by pressing the "control", v, and "alt" buttons at the same time.

5. <u>Delete duplicate rows:</u>
   - Again, sort the document, but this time by the "keep/delete" column. Do this by selecting the entirety of the data column, excluding the column title, and selecting "sort and filter" on the top right of the Home tab in Excel. Select "Sort z to A", press "Expand selection" in the dialog box, and press sort. This will sort all rows in the sheet by what is in the "keep/delete" column while preserving the corresponding data points in each row.
   - Select all rows at the top of the file that contain "move" in the "keep/delete" column. Copy these cells into another sheet that is labeled "[Registry name] "move"".
     - File rows are marked "move" when the Excel sheet finds a second address in a single foreclosure document. Since it is difficult to input multiple addresses for one foreclosure document into the database, we flag these documents, and don't include them in the final data file. However, we save them so that volunteers can later go through the documents and database and figure out how to include multiple addresses in the future.
     - It is important to note that this system flags the addresses that are the second or third address in a document as "move", so therefore the first document is still included in the data set. Thus, volunteers may need to manually connect flagged document IDs to document IDs that are not flagged and included in the final data file, to ensure that incomplete or incorrect data is not added to the database. This will need to be completed until the workbook can be improved to further streamline and automate this process.
   - Next, delete all rows that include "delete" in the keep/delete column. This condenses the data in the indexed data sheets into one row per document, with all grantor and grantee names in the one row.

6. <u>Clean and sort the data sheet:</u>
   - Sort the final data file by recorded date, street name, and street number by selecting "sort and filter" on the top right of the Home tab in Excel. Select "Custom Sort" and fill in the dialog box as shown below and press "OK".
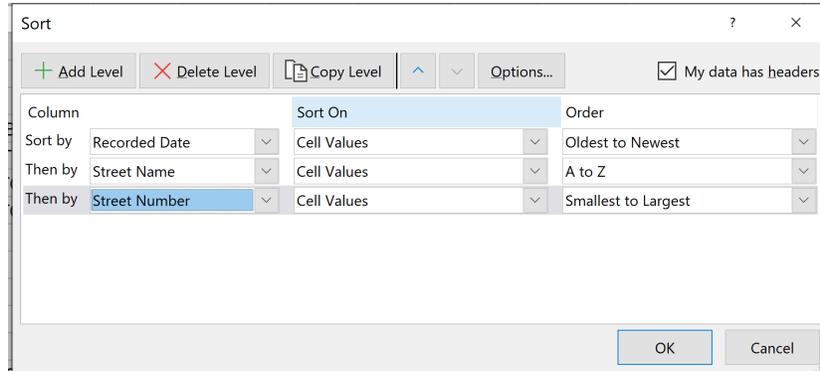
- Next, delete the "keep/delete" column from the final data by selecting that column, right clicking, and pressing "delete".
- Also, some Registries include "NULL" or "NONE" in cells when there is no data included. To better standardize this data, replace all instances of "NONE", "NULL", or "0" with an empty string: "". To do this, select the "Find and Select" button in the top right of the "Home" tab in Excel, and select "Replace". Enter the phrase you are replacing in the "Find what" field and what you are replacing it with, "", in the "Replace with" field. Also, expand the options by clicking the "Options" button, and select "Match entire cell contents". Make sure the settings look like the image below. Continue this for "NONE" and 0.
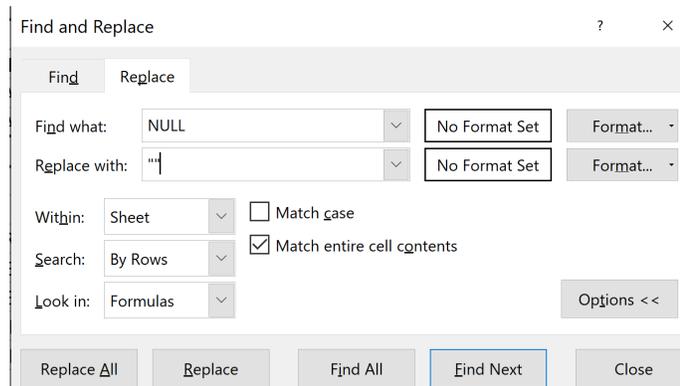
- Finally, select the entire Recorded Date column and change the data type included in that column to Date by right clicking and selecting "Format cells", "Date", and "OK". Format the Consideration column to "Number" type using a similar process.
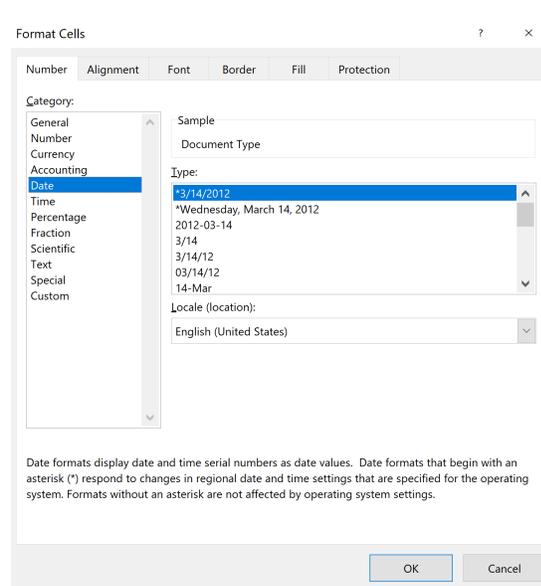
7. <u>Save the file:</u>
    - Finally, save the file as a CSV (comma-separated value) file for importing into the database. You only want to save the reformatted sheet as a CSV.

<u>Important Note:</u> Land Type

    Some registries designate registered and recorded land in their indexed data files by separating them into different sheets on the spreadsheet. However, the Berkshire Middle Registry does not contain anything to denote the land type directly. Instead, the indexed data provided includes both Registered and Recorded land. The book numbers with the lower two-digit books (20, 30, 40's) are Registered Land and the 4-digit book numbers are Recorded Land.

    In the Indexed Data Standardization Workbook, this needs to be denoted manually by the person inputting and reformatting the data. To do this, after step 3 of the procedure described above, sort the newly reformatted data by book number. In the Land Type column, denote the lower book numbers as "REG" and the higher book numbers as "REC".

Recommendations and Shortcomings:

    This system is not perfect. As described earlier in this document, the "move"/"delete"/"keep" designation flags data rows as potentially problematic due to documents including multiple addresses. However, this does not keep the first address and associated data from being included in the final data. This is a major bug that needs to be fixed before trusting this system completely.

    In addition, this system cannot reformat Power of Attorney files because the equations that consolidate the data into one line per foreclosure document and denote names as Grantor/Grantee 1, 2, 3, etc. rely on looping through the party types and party names in the original data. Because the program "looks" for Grantor names first in the list, if Grantee comes first, Grantor names will be

filled in after the Grantor 1 column. Thus, ideally there would be twice the number of maximum names for each Grantors and Grantees. Given there can be up to 20 names on a Power of Attorney, that would require 40 Grantor columns and 40 Grantee columns to avoid losing data. That is not sustainable for a file like this, so we strongly recommend against using this system for indexed Power of Attorney data to avoid losing data.

Because of this issue, we recommend improving the programming in this workbook to better accommodate data from all document types.

Characteristics of Data from Individual Counties

Each registry collects and outputs data in different forms. Below are some observations about the data received from the 5 registries we were able to contact.

Registry of Deeds Name	Notes
Dukes	- Must convert from Format A to Format B
Nantucket	- Must convert from Format A to Format B
Northern Middlesex	- No column headers in spreadsheet: A – date recorded B – document type C – consideration D – Grantor = D; Grantee = I E – Document number F – LastName G – FirstName H – Street Number I – Street Name J – Town K – Book L – Page - Different document type codes for registered and recorded land: FD = recorded land foreclosure deed MTG S&A = registered land foreclosure deed PA = recorded land power of attorney P ATTY = registered land power of attorney
Southern Essex	- Registered land has no book or page number - No grantor/grantee (party type) designation - Recorded land contains abbreviated town names - Document type abbreviations: POSN = Possession (Certificate of Entry), DPO = Foreclosure Deed POW = Power of Attorney

Middle Berkshire	- Indexed data provided includes both Registered and Recorded land. If you sort by book number the lower two-digit books (20, 30, 40's) will be Registered Land and the 4 digit book numbers will be Recorded Land.
Northern Bristol	- RL: Recorded land - LC: Registered land - "POSSESSION": Certificate of Entry - Includes power of attorney (remove from final data) and affidavit of sale (include in final data)
Suffolk	- POA: power of attorney - EOP: evidence of possession - FDD: foreclosure deed Extracted data includes: - POA's that have EOP and FDD as next 2 doc numbers - EOP that have FDD as next document number - FDD that have EOP as document number before it - Foreclosure deed was misspelled "Forclosure Deed" in the data we received