Developing a Data System to Address Foreclosures in Massachusetts

Cape Town Project Center 2020

More at https://wp.wpi.edu/southafrica/projects/2020-projects/data/

Abstract

This project assisted the Massachusetts Alliance Against Predatory Lending (MAAPL) in addressing the ongoing home foreclosure crisis in Massachusetts. This was done by gathering data into an organized Data Collection and Management System to analyze trends in wrongful foreclosures in the state. The Data System includes a set of Guides for Public Data Collection, a functioning database created with SQL code, and data that has been gathered from public sources and imported using Python scripts. We also developed a guide for the future development of the system. The project also helped to develop an initial relationship between WPI and MAAPL and built the foundation for future IQP teams to continue this work.

Team Members

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

Project Advisors

Prof. Scott Jiusto Phd
Prof. Gbetonmasse Somasse Phd

Project Sponsor & Liaison

Ms. Grace Ross
Massachusetts Alliance Against Predatory Lending (MAAPL)
and Worcester Anti-Foreclosure Team (WAFT)

Table of Contents

Executive Summary

Foreclosures and evictions have become a major crisis in America. During the beginning of the housing crisis, between 2007 and 2009, there were 2.5 million foreclosures nationwide (Aguirre & Martinez, 2014). Today's trends parallel those during the earlier part of the millennium, with more than 300,000 Massachusetts residents unable to pay their rents or mortgages in July 2020 and over one quarter of the country in the same situation. This current rise in foreclosures is a direct result of the economic downturn caused by the COVID-19 pandemic, as well as lingering effects of the housing crisis and other economic downturns (Healy & Rios, 2020).

The process of foreclosure is lengthy and complex. Foreclosures occur when homeowners can no longer afford their monthly home mortgage loan repayments. Homeowners have the right to advocate for themselves and fight their foreclosure and eviction cases in court if they believe that they have been wrongly foreclosed on. However, since the foreclosure process is complicated and involves many court dates and official documents, homeowners often do not understand their rights. This complexity also makes it simple for banks and mortgage companies to skip steps of the foreclosure process and take advantage of unknowledgeable homeowners (MAAPL, 2020). Therefore, if homeowners are not knowledgeable of the process or do not have experts advocating for them in court, they can lose their home when there may have been other options available to them.

One of the ways in which banks and mortgage companies have targeted homeowners is through the predatory lending practice of granting subprime mortgages, which is the main practice that sparked the 2008 financial crisis, and still takes a toll today. A subprime mortgage has either a very high initial interest rate or an interest rate that increases over the life of the loan. This means that the mandatory minimum payment will increase as time goes on, making it difficult for many to complete payments, especially if a borrower's income declines. Because of the current COVID-19 pandemic, with a plummeting economy and increasing unemployment, foreclosure and eviction pressures are rising rapidly for homeowners with both subprime and more legitimate mortgages (Jurow, 2020). Several Worcester organizations have been working to address this growing concern.

The sponsors of this project are the Massachusetts Alliance Against Predatory Lending (MAAPL) and the Worcester Anti-Foreclosure Team (WAFT). MAAPL, our primary sponsor, fights foreclosures by guiding homeowners through the foreclosure process and advocating for legislation to protect homeowners.

MAAPL and WAFT have been successful in helping many homeowners stay in their homes but have a need for data that would allow them to make large-scale legislative change. MAAPL is aware of several instances within the foreclosure process in which lenders commonly cut corners, skip steps, and perform illegal actions, but to instigate legislative change or collective legal action within the state of Massachusetts, they need to analyze large amounts of data if they are to prove these illegalities are common practice among significant numbers of banks and mortgage companies. The data that MAAPL needs is all

publicly available in thousands of documents held by the Registries of Deeds for Massachusetts counties and within the US Securities and Exchange Commission. However, each foreclosure involves over 5 documents and each of the 21 Registries of Deeds in Massachusetts records dozens, if not hundreds, of foreclosures per year. A full set of these foreclosure documents is required to fully understand an individual foreclosure, and therefore, identifying patterns in foreclosures throughout

Figure 1 Excerpt of Data System ERD

Massachusetts requires linking the data in many separate documents. In addition, the Registries of Deeds all use different website formats, data management systems, and data conventions, making it difficult to generalize the data collection process. The legal documents are also only available as image files as opposed to machine readable text, and no one else has gathered the information in a central repository.

This project resulted in three major outcomes to address the problem described above,

1. **Data Collection and Management System** ("Data System"): The main deliverable is a Data System that can store data collected either manually by volunteers, by using automated processes for data scraping, or received in batches from cooperative Registries of Deeds and others. The system contains
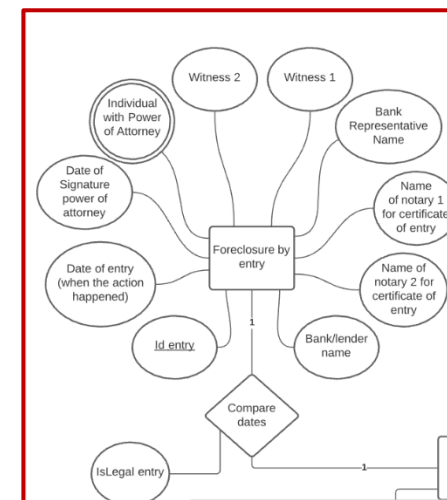
a relational database which allows for data values collected from various legal documents to be organized and related in several tables (Figure 1), from which users can query for specific information related to a foreclosure. It also includes a repository of data that has not been entered into the database and methods for standardizing and inputting the data collected from different registries. This information can be used to analyze illegalities in the foreclosure documents on a large scale, while also providing tools to assist in fighting individual foreclosure cases, which is a major goal for the sponsor.

2. **Guides for Public Data Collection**: There are many volunteers who support MAAPL and would be able to input data from legal documents into the database. Because the government websites where this data is located are difficult to navigate, we created Guides for Public Data Collection to help volunteers collect data from the websites for each Massachusetts County's Registry of Deeds and the US Securities and Exchange commission.

3. **Future Development Guide**: After the conclusion of this project, there will still be many opportunities to expand upon the team's work and further develop the Data System. Because of this, the team has created a Future Development Guide outlining the current functionality of the system, gaps within it that should be addressed, recommendations for data analyses to take on, and other information for future project teams, MAAPL, and others.

Background
The Ongoing Foreclosure Crisis in Massachusetts

According to the US Department of Housing and Urban Development, "for many Americans, owning a home is an essential part of the American dream that conveys a number of economic benefits, such as the ability to accumulate wealth and access credit by building home equity, reduce housing costs through the mortgage interest deduction, and gain long-term savings over the cost of renting" (n.d.). Homeownership is an important part of society, but foreclosures, which result in people losing their homes, reduce these benefits for many people. Foreclosures can have a multitude of causes and can be escalated by major global events. For example, massive waves of foreclosures have led up to and followed events such as the 2008 Housing Crisis and the COVID-19 pandemic (Davis, 2008). Both events put a strain on the job market, leading to an increase in foreclosures across the country.

The process of foreclosure begins when a borrower fails to pay their mortgage payment for at least three months (MAAPL, 2020) This inability to pay can occur due to a variety of reasons. Under normal conditions, defaulting can be caused by catastrophic unforeseen events such as death, job loss, illness, or divorce. These events often have a substantial effect on household income but may still be overcome if the mortgage rate is sustainable. In addition to these causes, discriminatory lending practices result in loans being offered at unsustainable rates to certain populations. This leads to increased foreclosure rates in some areas, especially where many minorities live.

After a borrower fails to pay his or her mortgage payment, there are a series of steps that must be followed

for a bank or mortgage company to legally foreclose on a home, which are different in each state. A property can be foreclosed on with two different methods. The foreclosing party may enter the property peaceably and unopposed in "foreclosure by entry", or the property may be auctioned off in "foreclosure by auction". Additionally, in Massachusetts specifically, the process is non-judicial, meaning that a court case is not required to foreclose on a home (MAAPL, (n.d.)). Because of limited judicial oversight, it is easier for banks to take advantage of homeowners and not follow the legal foreclosure process correctly. The foreclosure process in Massachusetts is outlined in Figure 2 below.
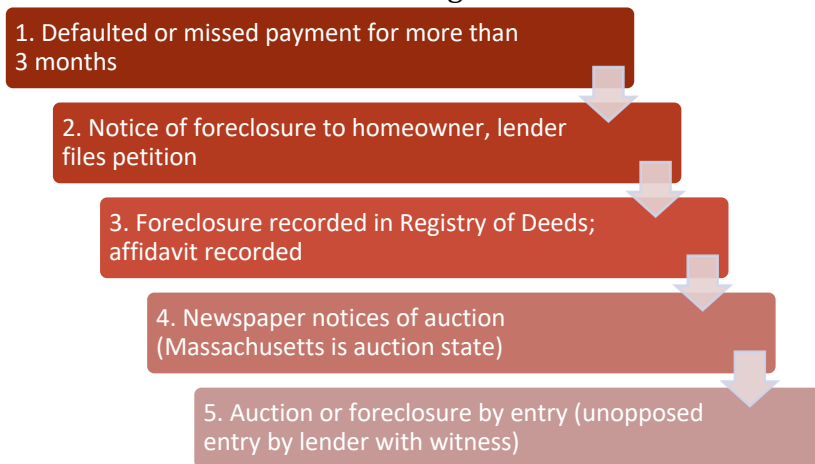

1. Defaulted or missed payment for more than 3 months

2. Notice of foreclosure to homeowner, lender files petition

3. Foreclosure recorded in Registry of Deeds; affidavit recorded

4. Newspaper notices of auction (Massachusetts is auction state)

5. Auction or foreclosure by entry (unopposed entry by lender with witness)

Figure 2 Foreclosure process steps

The complexity of the foreclosure process also leaves a lot of room for systemic illegalities and predatory practices. If the lender follows the proper procedures exactly then the foreclosure process protects the homeowner. However, the inherent complexity in the foreclosure process makes it difficult for the homeowner to tell if the lender is following the process correctly. As a result, there are many opportunities for a lender to take advantage of a borrower, such as by skipping steps or not following the foreclosure procedures exactly. For example, at both foreclosures by entry and auction, an authorized bank representative must be present to complete the foreclosure. However, often the bank representative is not legally authorized at the time of the foreclosure (Ross, 2020). Small missteps like this and many others wrongfully take advantage of homeowners, who are often not knowledgeable of the process, and benefit big lenders and auctioneers by expediting and simplifying the process. Homeowners must know the prevalence of such problems so they can identify and contest or negotiate over wrongful foreclosures. This makes reaching out to them to inform them of their rights a very important step in fighting wrongful foreclosures statewide.

Addressing Foreclosures and Evictions in Massachusetts

Due to the problems described above, members of the community have found many different ways to help. Groups such as the Massachusetts Alliance Against Predatory Lending (MAAPL) and The Worcester Anti-Foreclosure Team (WAFT) were formed to support individuals in the community trying to save their homes from foreclosure

The Worcester Anti-Foreclosure Team (WAFT) is a local grassroots organization aimed at providing mutual aid and information to homeowners facing foreclosure. WAFT was founded in the spring of 2008 (Khan, 2016) in the wake of the 2008 financial crisis. The group's main objectives are to educate homeowners on their foreclosure status by hosting monthly meetings, door knocking, and distributing brochures. WAFT also aids homeowners in

maneuvering through the legal processes of foreclosure (Ross, 2020). Most of their efforts are focused on minority communities that are targeted disproportionately for subprime mortgages.

Figure 3 WAFT members celebrating a stopped auction of foreclosed home (WAFT, 2016)

Similar to WAFT, the Massachusetts Alliance Against Predatory Lending (MAAPL) provides aid to homeowners, but throughout Massachusetts. MAAPL partners with other local organizations such as WAFT and provides resources in their fight against wrongful foreclosures. MAAPL also uses information from legal documents to prove that there has been systemic wrongdoing in the process of foreclosure in Massachusetts and advocates for legislation to end predatory lending and protect homeowners. Having an effective method to collect data from these documents and analyze the data will help MAAPL fight wrongful foreclosures more efficiently by providing proof of these widespread issues.

Public Data Collection and Analysis

By collecting data sets from several publicly available locations, organizing them into an analyzable form, and demonstrating trends in illegal activities by banks, MAAPL can create cases against these wrongful foreclosures and work to prevent more from occurring in the future.

There is a great volume of data that is available for the public to access related to foreclosures and evictions in Massachusetts. However, accessing the data is not a straightforward process. For an individual to find a deed, a mortgage, or any other documents related to foreclosures from the Registry of Deeds, the person needs to search for the document on the website for the county where the deed or foreclosure report would be located. Each Registry organizes its website differently, and often some specific information relating to the report or deed, such as the location of the document in the public record or the address associated with the document, must already be known (Glavin, n.d.). In addition, even when other sources and companies attempt to compile and report this data in one place, the data is often incomplete. These issues make it extremely difficult for organizations such as MAAPL to gather foreclosure documents in bulk to provide evidence of any systemic wrongdoing.

Another source of home mortgage data is the United States Securities and Exchange Commission (SEC) database of public documents. One document that is stored in this database and is of interest to MAAPL is the Pooling and Servicing Agreement. It is difficult to collect information from this kind of document because they are often difficult to find in the SEC database due to incomplete labeling and document searching capabilities. As a result, developing a

system to automatically collect information from these documents is extremely difficult.

Many kinds of publicly available documents describe foreclosure proceedings and are available as pictures on government websites such as each county's Registry of Deeds. This data is often key to court cases against wrongful foreclosures and showing systemic trends in foreclosures. Using computers to quickly collect data from documents like these is ideal, but for computers to be able to understand the words on these pages they need to use a technology called Optical Character Recognition, or OCR. OCR is a system that converts an image of a text which is machine-printed into machine-readable text (Cheriet et al., 2007). After scraping the relevant data, it is important to "clean" the data. The process of data cleaning ensures that the data contains no errors or irregularities. As shown in Figure 4, the data cleaning process can include reformatting the data to best fit the structure needed for analysis. Some aspects of reformatting that may be relevant to this project include standardizing names of borrowers and banks involved in foreclosure cases, reformatting addresses to distinguish between the street address and apartment numbers, removing any unexpected duplicates that may be present in the data, and geocoding addresses (Desmond, 2018).

Massachusetts policymakers and organizations fighting foreclosure in Massachusetts would benefit greatly from a user-friendly Data System which incorporates systematic data collection from the Registry of Deeds and other sources, and the input, organization, and analysis of that data. This would alleviate many of the drawbacks of the publicly available data by accessing and consolidating a large quantity of data into one place and would make it

easier for organizations like MAAPL to discover wrongful foreclosures and predatory lending. With this consolidated data, MAAPL and other organizations could present evidence of systemic wrongdoing to compete against large corporations and advocate for major systemic and policy change. It could also potentially benefit individuals in the foreclosure or eviction processes to determine if they were the victim of wrongdoing.

DATA CLEANING

1 - Screen	2 - Diagnose	3 - Treat	4 - Document
Locate and identify errors, inconsistencies, and outliers in the data	Determine the specific problem with these errors	Decide whether to change, leave, or delete the inconsistent data	Keep track of changes made and store original data before changes

Figure 4 Steps for data cleaning: Adapted from (The Assessment Capacities Project, 2016)

Methodology: Designing and Implementing MAAPL's Data System

The methodology for this project worked to achieve the following objectives (Figure 5). In addition to the methodology described in this section, the team followed a detailed methodology that is further explained in the Future Development Guide.

Figure 5 Data System Methodology, adapted from Taylor (2014)

Identifying Data Needs and the Vision for the System

The team first identified the state of MAAPL's existing data organization and collection systems and how they could be improved through interviews and email communications with the MAAPL liaison, Grace Ross. In addition to informing the team, these interviews and meetings also ensured that the final Data System would help MAAPL achieve its own goals. The team also determined what data and information were the most important to include in the system to address MAAPL's highest priority analysis goals: proving trends in illegal actions and seeking redress through lawsuits and policy change.

Designing the Data Collection and Storage System

The team planned the organization of the system and attempted to implement it following the plan that was developed through interviews with MAAPL. The process was meant to match the system to the vision as effectively as possible.

Planning Data Organization

The first step in the Data System design process was understanding the components of the data system, including data sources, data points to be collected, and data cleaning. This was completed by creating a flow chart to visualize these components of the system and how they connect.

The team also accessed all the possible data sources to determine what data can be collected from each source and the methods that are required to create collections of that data. Using this information, the team built a detailed Excel sheet describing the data fields to be collected, from what sources they are collected, and how they are collected and input into the database. By doing this, the team was able to organize their ideas about the overall flow of the Data System and what would be required to collect all the data points that MAAPL wanted.

In addition to understanding the process, it was helpful to understand how these data points all related to each other within the data system. This part of the Data System planning was completed by creating an Entity Relationship Diagram (ERD). An ERD organizes the entities, attributes, and relationships in the Data System to visualize how the data collected in the system relate to each other. Creating the ERD helped the team plan the individual data values that will be collected within the data system, as well

as how the data are connected, to create a robust relational database.

Implementing and Populating the Data System

With the Data System outlined and planned out, final technical steps were followed to build and populate the data system. These important steps are outlined below, with further details for each step outlined within the Future Development Guide:

Step 1: Data Collection
Step 2: Create Data Input Protocol
Step 3: Clean the data
Step 4: Establish data connections
Step 5: Iterative Design

Planning for System Sustainability and Data Privacy

To keep our sponsor involved in the entire process of the project, the team had to consider their involvement both in the design of the system and the planning of the future of the data system. This included continuously communicating with MAAPL throughout the process of the project to make their needs were being met and adapting as those needs changed. Including MAAPL in every step of the process and creating detailed documentation for the system set up our Data System for long-term success.

An important part of these communication and documentation efforts was creating guides to inform MAAPL volunteers and future project groups on the workings of the data system, any recommendations for improving the system, and instructions for how to collect and reformat data to maintain the system. These guides include:

1. Future Development Guide
2. Guides for Public Data Collection
3. Data Reformatting Guide and Excel Workbook

Data Privacy and Research Ethics Concerns: the team worked together with the WPI Institutional Review Board to ensure that the project was ethical and to address any issues that future teams may encounter when continuing our work.

Project Findings and Outcomes

The Data System created during this project includes two major components: the Data collected from the Registry of Deeds and the Database Code which stores and processes data into tables (Figure 6). The Registry of Deeds data is reformatted and standardized using a master Excel workbook. SQL code is used to create the data tables and query data from them, and a Python script populates the database with the Registry of Deeds data. Additional outcomes include guides to help collect and reformat data, use the database, and guide future development of the data system. These supplemental materials can be located on this project's webpage on the Cape Town Project Center (CTPC) website.
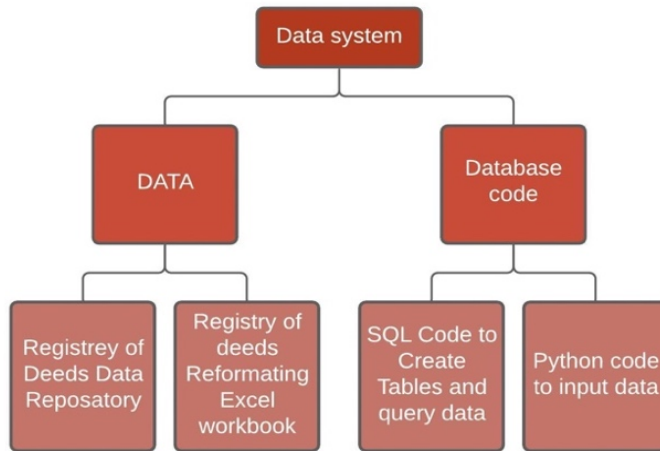
Figure 6 Data System structure

Data System Planning & Implementation

From interviews and meetings with MAAPL, the team determined that MAAPL wished to have the data to investigate the following data analyses of interest:

1. Analysis 1: Investigate whether mortgage companies intentionally provided individuals with mortgages that exceeded the value of their homes.
2. Analysis 2: Investigate the legality of pooling mortgages.
3. Analysis 3: Investigate illegalities and inconsistencies in the foreclosure process, especially around home auctions, in hopes of identifying systemic patterns that could halt foreclosures and evictions and end the most prevalent illegal foreclosure practices.

Completing these analyses requires many data points from different sources using multiple methods. (Table 1).

The team also collected data from 7 out of the 21 Registries of Deeds throughout Massachusetts in the form of electronic images of documents as well as recorded data that was available in spreadsheet format and therefore more easily manipulated and added to the database. The Registry of Deeds system in Massachusetts is decentralized and individual counties and regions have separate systems to record documents and the electronic indexed data was received in many different forms. Therefore, these data files had to be standardized using formulas and programs in Excel to input the data into the database.

The team's Entity Relationship Diagram (ERD) reflected and visualized the planned connections between the data in the database. This diagram (Figure 7), which was created using the online visualization software Lucidchart, allows for a simpler transition from planning to coding and will make querying the database more straightforward for a user. In addition, having all the data points related to a specific legal process, such as foreclosure by entry or foreclosure by auction, allows for easier comparisons of data and identification of illegalities in these legal processes.

The ERD was then translated to SQL code that specifies the tables and relationships shown in the ERD and present in the final database. The code was used in parallel with a MySQL relational database to input and query data from the database. Additional code was also written in Python to transfer data stored in Comma Separated Values (CSV) into the database and can be viewed in the supplemental materials.
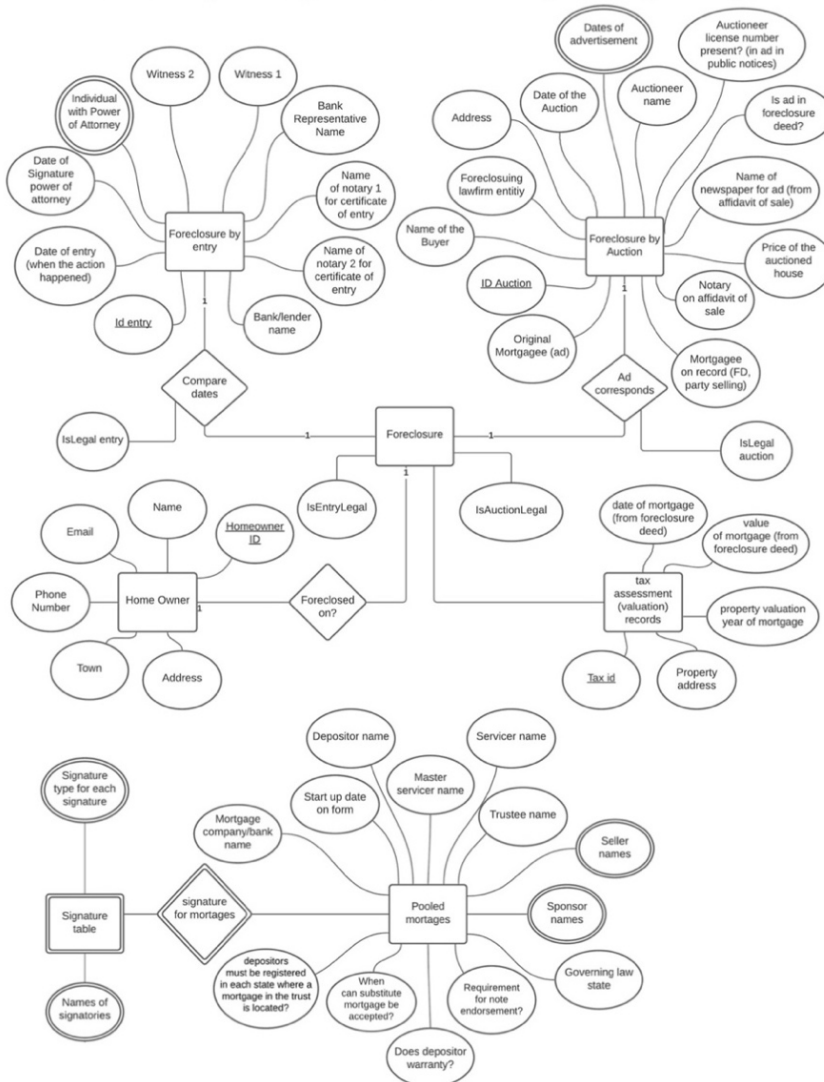
ERD (Entity-Relationship Diagram)

Figure 7 Database Entity-Relationship Diagram (ERD) showing the contents of the Database

Future Development Guide

Another major outcome was creating a Future Development Guide to help educate future users and developers of the Data System on how to use and improve the current system. This guide includes a technical explanation of the Data System and justification of our methods to assist others in understanding our work. We assessed which groups could complete certain aspects of the Data System and made preliminary plans and recommendations for these groups in this guide. Important topics covered in the Future Development Guide are listed below.

1. **How to use the data system**
   - How to rebuild database using foundational files stored on WPI server
   - How to collect data to populate the system, including contacting different data sources, accessing data, and extracting data
   - How to input data into the system, such as more details on reformatting files received from Registries of Deeds, and adding data into the system using spreadsheets
   - Rules and conventions established for the Data System
2. **Recommendations and future projects**
   - This topic is discussed in the section titled Recommendations for Future Work

Guides for Data Collection and Input

Finally, the team created guides to help with handling data in the future. Throughout the process of planning and implementing this data system, the team documented various processes for MAAPL volunteers and

future project teams to be able to reference. This was important for the database to be used and developed in the future. In addition, these guides were helpful to MAAPL to train the organization's many volunteers in important repeated processes of collecting and inputting data. Guides that were created based on the team's work include:

1. Guides for Public Data Collection
2. Guide for Reformatting Indexed Data from Registries of Deeds

The Instructional Guides for collecting data from the Registries of Deeds and SEC were created for MAAPL volunteers to use when adding data to the database. As more foreclosures occur, MAAPL will need to continuously add new information to the database. These guides outline how volunteers should navigate the websites for each county's Registry of Deeds to efficiently collect desired data points from each document type. Because there are multiple different Registry website layouts, there are Instructional Guides for each unique website layout.

The guide for reformatting indexed data outlines the methods the team took to standardize the formatting of the indexed CSVs sent by the various county's Registers. Creating a standard format for the data within the CSV files between the counties simplifies the process for inputting the data from the CSVs into the Data System and makes it easier to train volunteers to complete this task. This standardization will minimize the errors encountered when a program runs through the process of transferring the data into the system and limits the amount of human interaction needed for that process.

Recommendations for Future Work

This project was just the beginning of automating MAAPL's data collection, storage, and analysis processes and presents many interesting opportunities for future projects with WPI and other MAAPL associates. We have compiled several recommendations to guide the future development of the data system. These suggestions could be pursued by the WPI London Project Center IQP team that will be continuing our work next semester or other teams or volunteers.

1. **Automate the data collection and input processes**

   MAAPL would benefit greatly from a program that automatically collects and inputs data into the database when data sources such as the Registry of Deeds and Massachusetts Public Records are updated.

   The system could also be further automated and streamlined by developing a program to OCR handwritten fields in the Registry of Deeds documents, which would save MAAPL's volunteers a lot of time and effort and reduce human error when inputting data.

2. **Analyze and visualize the data**

   Once all data is collected, the team has made suggestions for analysis and visualization. These analyses should aim to fulfil the data analyses goals described previously.

   An example of a worthwhile analysis to perform is comparing the dates of the foreclosure by entry to the date of when the bank representative at the foreclosure was authorized to act as a representative through a power of attorney. If the representative was not authorized at the time of the foreclosure, then that

suggests the foreclosure was legally suspect. There are many other examples of potential illegalities found in Registry of Deeds documents that can be analyzed once the database is fully populated.

3. Refine MAAPL's data intake form

A major request of MAAPL's was to build out their current data intake form, which asks at-risk homeowners many questions about their foreclosure case and automatically creates important legal forms using this information to help the homeowner fight their foreclosure case in court. The data from this intake form could also be input into the database to help MAAPL track foreclosures in real time and retain contact information and key information for the individual's case. The team was not able to address the issues with this system during the project term, but this would be an important task for future teams to complete as it will greatly streamline MAAPL's case intake processes and help homeowners fight their cases in court.

This project was reviewed and approved by the Worcester Polytechnic Institute Institutional Review Board (IRB). The team initially proposed completing the data intake form section of the Data System during this project term, but quickly decided that was outside of the project's scope. However, the team still worked together with the Institutional Review Board (IRB) at WPI to create a consent form since this system deals with personal data that is not publicly available. We suggest using that form as a draft for a consent form when implementing the data intake form. This form, as well as other recommendations for ensuring data

privacy, is included in the supplemental materials available to teams and individuals continuing our work.

4. Create a case management system:

The collected data could be used to create a dashboard of information relating to foreclosures in Massachusetts and specific individual cases. These dashboards would help inform MAAPL volunteers and potentially members of the public of the overall state of foreclosures in Massachusetts. It would also help MAAPL and its constituent local organizations better assist at-risk homeowners by presenting important information relating to each individual foreclosure that will be helpful in tracking and fighting each foreclosure case.

5. Recruit WPI groups and others to finish technical aspects of the system:

Finally, many of these analytical and development tasks will require more time and skills than are available at MAAPL. Therefore, we recommend that WPI and MAAPL create a sustained partnership in which IQP teams, MQP teams, and research groups work to tackle MAAPL's developing system and technology needs. In addition, the organization may consider outsourcing code that has especially time sensitive deadlines. The team has started pursuing some of these options already, but others are recommendations for future groups. These options are shown in Table 2 and explained further in the Future Development Guide.

Conclusions and Project Impact

The team helped MAAPL lay the initial foundation for a data collection, storage, and analysis system to fight illegal foreclosures more efficiently and effectively in Massachusetts.

The data collection process from the Registry of Deeds documents is very intensive when done manually. Automating this process with a Data System that scans for important information, cleans and reformats it, and presents it in a useful way for various audiences greatly speeds up MAAPL's ability to advocate for policy that protects borrowers and homeowners. Being able to prove the systemic nature of illegal actions in foreclosures in Massachusetts will make a lasting impact on the ongoing foreclosure crisis.

In addition, starting an ongoing partnership between WPI and MAAPL is beneficial for both organizations, providing WPI students opportunities to make real change in Worcester and helping MAAPL make their processes more efficient and effective.

Acknowledgements

We would like to extend our sincere gratitude to the following people who were instrumental to this project.

To our sponsor, Grace Ross, MAAPL's founder and foreclosure expert, for the guidance, expertise, and passion which drove our project.

To our advisors, Gbeton and Scott, for always pushing for excellence and their time in helping this project be successful.

To Steve Floridia for his expertise in data systems and foreclosures and invaluable assistance in helping us get our system off the ground. Without you we never would have gotten this far!

To the many Registries of Deeds in Massachusetts and the many individuals who gave their time in helping us gather data for our project.

Finally, to all our classmates and the Outreach Team, who made this term a lot more enjoyable.

Useful Tables

Table 1: Data Sources of Interest

Sources	Data Available	Data Collection Method
Massachusetts Registries of Deeds	- **Foreclosure Documentation** - **Property History** - **Indexed data and scanned documents**	- **Indexed data from individual registries** - **OCR**
US Securities & Exchange Commission	- **Found national documentation for bundled mortgages and securitized trusts**	- **OCR** - **Text scraping script**
Massachusetts Public Notices	- **Newspaper ads for foreclosure auctions** - **Foreclosure details**	- **Text/web scraping script**
CoreLogic/MLS Real Estate database	- **Property sales, foreclosure, and tax assessment history** - **Property details** - **Owner demographics**	- **Text/web scraping script** - **Some data may come in spreadsheets, but unknown**

Table 2: Future Database Development

Options for Continuing Database Development
Future MQP/IQP Projects (ongoing)
Partner with WPI professors or research groups doing anything similar or looking for new projects (ongoing)
Hackathon challenge to create one of the aspects that are high priority for MAAPL
Outsource coding work to finish high priority projects quickly and correctly

References

Aguirre, A., & Martinez, R. O. (2014). The Foreclosure Crisis, the American Dream, and Minority Households in the United States: A Descriptive Profile. *Social Justice*, *40*(3 (133)), 6–15. JSTOR.

Berry, S. H., Burger, N., Dogo, H., Kumar, K. B., Malchiodi, A., Martini, J., Mengistu, T., Shatz, H. J., Smith, A. C., Usanov, A., & Yoong, J. K. (2014). Methodologies for Data Collection and Handling. In *Designing a System for Collecting Policy-Relevant Data for the Kurdistan Region—Iraq* (pp. 33–52). RAND Corporation; JSTOR. https://www.jstor.org/stable/10.7249/j.ctt6wq9bd.13

Cheriet, M., Kharma, N., Liu, C., & Suen, C. (2007). *Character Recognition Systems* (1st ed.). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470176535

Christensson, P. (n.d.). *OCR (Optical Character Recognition) Definition*. Retrieved October 1, 2020, from https://techterms.com/definition/ocr

Cornillon, P. (n.d.). *What Is a Data System, Anyway? | EDUCAUSE*. Retrieved September 23, 2020, from https://er.educause.edu/articles/2005/1/what-is-a-data-system-anyway

Dane, M. (2018, April 16). *SQL Tutorial—Full Database Course for Beginners*. https://www.youtube.com/watch?v=HXV3zeQKqGY

Department of Housing and Urban Development. (n.d.). *Homeownership: The American Dream | HUD USER*. Retrieved October 1, 2020, from https://www.huduser.gov/portal/pdredge/pdr-edge-frm-asst-sec-081318.html

Desmond, M., Gromis, A., Edmonds, L., Hendrickson, J., Krywokulski, K., Leung, L., & Porton, A. (2018). *Eviction Lab Methodology Report Version 1.0*. Princeton University Eviction Lab. https://evictionlab.org/docs/Eviction%20Lab%20Methodology%20Report.pdf

Effective Database Managment. (n.d.). *Five Steps to Ensure the Success of Your Database – Effective Database Management*. Retrieved September 30, 2020, from https://effectivedatabase.com/five-steps-to-ensure-the-success-of-your-database/

Glavin, W. (n.d.). *Welcome to 20/20 Perfect Vision Land Records I2*. Retrieved September 23, 2020, from https://www.masslandrecords.com/Worcester/

Healy, B., & Rios, S. (2020, August 7). *Housing Crisis Looms As Mass. Renters And Homeowners Miss Payments*. WBUR News. https://www.wbur.org/news/2020/08/07/housing-eviction-mortgage-massachusetts-renters-homeowners-coronavirus

Khan, A. (n.d.-a). *What does WAFT do, anyway? An inside look at Worcester's anti-foreclosure warriors: Part 1 — Gaining traction and attention*. Worcester.Ma. Retrieved September 24, 2020, from

https://worcester.ma/2016/09/what-does-waft-do-anyway-an-inside-look-at-worcesters-anti-foreclosure-warriors-part-1-gaining-traction-and-attention/

Khan, A. (n.d.-b). *What does WAFT do, anyway? An inside look at Worcester's anti-foreclosure warriors: Part 2—The team rallies around an Oak Street family | worcester.ma*. Retrieved September 24, 2020, from https://worcester.ma/2016/10/inside-look-at-worcesters-anti-foreclosure-warriors-part-2-team-rallies-around-oak-street-family/

La Torre, J. (n.d.). *6 Steps for Best Practices in Database Design—DZone Database*. Dzone.Com. Retrieved September 30, 2020, from https://dzone.com/articles/what-are-the-best-practices-of-database-design

Lab, E. (n.d.). *The Eviction Lab*. Eviction Lab. Retrieved September 24, 2020, from https://evictionlab.org/

Lucidchart. (n.d.). *What is an Entity Relationship Diagram (ERD)?* Lucidchart. Retrieved October 29, 2020, from https://www.lucidchart.com/pages/er-diagrams

Lucidchart. (2018, April 16). *Database Tutorial for Beginners*. https://www.youtube.com/watch?v=wR0jg0eQsZA

Massachusetts Alliance Against Predatory Lending. (2012, May 24). *MAAPL History*. MAAPL – The Massachusetts Alliance Against Predatory Lending. http://maapl.info/about/history/

Microsoft. (n.d.). *Database design basics*. Retrieved September 30, 2020, from https://support.microsoft.com/en-us/office/database-design-basics-eb2159cf-1e30-401a-8084-bd4f9c9ca1f5

Reeb, B. (n.d.). *8 Steps to Build a Data Systems Inventory*. Retrieved September 29, 2020, from https://blog.idatainc.com/8-steps-to-build-a-data-systems-inventory

Registry of deeds. (n.d.). *Barnstable County Registry of Deeds |*. Retrieved September 23, 2020, from https://www.barnstabledeeds.org/

Soffer, A. (2019, July 10). 7 Best Practices for Effective Data Management in 2019. *Leadspace*. https://www.leadspace.com/best-practices-for-effective-data-management/

Taylor, R. (2014). The Framework for Data Systems. The Institute of Educational Sciences (IES) Statewide Longitudinal Data Systems (SLDS) Grant Program. https://nces.ed.gov/programs/slds/pdf/framework_for_data_systems_may2014.pdf

Vanderstreet, J., Eusman, N., Boyd, S., & Kalimani, S. (n.d.). Design of a Database Platform for Eilat. 54.

Weber, A. (n.d.). *14 No-Fail Steps to Building a Database*. Target Marketing. Retrieved September 30, 2020, from https://www.targetmarketingmag.com/article/14-no-fail-steps-building-database-28255/

Zelic, F., & Sable, A. (2020, July 31). *[Tutorial] OCR in Python with Tesseract, OpenCV and Pytesseract*. AI & Machine Learning Blog. https://nanonets.com/blog/ocr-with-tesseract/\