ASSOCIATION OF CANADIAN MAP LIBRARIES AND ARCHIVES

BULLETIN

Visualizing Kitchener: Geocoding Historical Street Information

Eva Dodsworth, *Geospatial Data Services Librarian*
Nolan Pilozo-Hibbit, *Geomatics Co-op Student*
Markus Wieland, *GIS Specialist*

University of Waterloo

Abstract

When studying and investigating the history of a property or neighbourhood, common resources often sought by researchers include air photos, fire insurance plans and historical maps. These will provide information about what an area looked like, often offering the building's footprint and surrounding landscape. Timespan studies will no doubt show growth, development and possibly changes to the buildings of interest. But what the rich resources don't tell the researcher is information about the people connected to those buildings. Who lived or worked there? What did they do for a living? Did they move often? Did they change jobs regularly? These types of questions can't be answered with just maps alone as they require a detailed census to go along with it. To fill this type of need, Geospatial Centre staff at the University of Waterloo Library embarked on a massive-scale digitization and geo-location city directory project – one that had started in 2019, has involved at least 40 staff members, and still has a couple years to go before completion. This paper will summarize the project thus far, with a focus on the journey of geocoding historical streets.

Funding and Acknowledgement

This project has been funded in part by the Social Sciences and Humanities Research Council, as well as the University of Waterloo. We would like to acknowledge support from Karen Ball-Pyatt, from the Kitchener Public Library, Darryl Bonk, from Waterloo Region Generations, Waterloo Regional Heritage Foundation, and the Waterloo Region Branch Ontario Genealogical Society who have generously provided us with scanned versions of the city directories.

Introduction

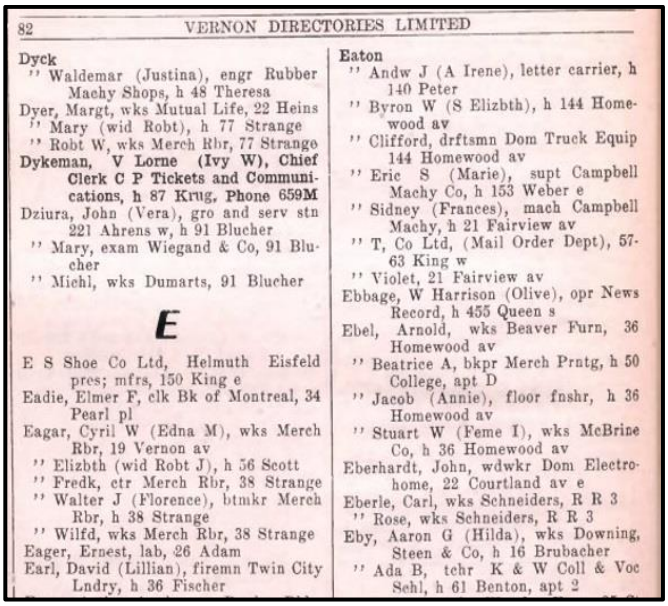
The Geospatial Centre, University of Waterloo has been working on a significant data transformation and visualization project aimed at providing scholars, researchers, and the public with a visual directory, index and account of Kitchener, Ontario's residential, business and industrial development. As one of the most detailed historical records of every building, household

and business, the Vernon's city directories for Kitchener-Waterloo, published by Henry Vernon & Sons, offer narratives, facts, insight, and clues about life in Berlin/Kitchener throughout the years (Kitchener was named Berlin up until 1916). Whether it is to study family history, business retention, impacts of war and/or industrialization, extracting and analyzing information from these historical gems will no doubt aid in the discovery and learning process. Using the spatial information, coupled with the business advertisements, researchers can visualize all the moving parts together, creating opportunities for new connections.

What would traditionally take painstaking work, sifting through and analyzing individual volumes (one per year), and entries of the print resource, Geospatial Centre staff aimed to clean up, collate and geocode the city directory information, offering the results in a searchable online map, providing not only digital access to records from multiple volumes of the directories, as well as search, and filter features, but more importantly, the ability to visualize these hundreds of thousands of records, offering users an understanding of *where* people resided, *where* they worked, and the ability to discover spatial connections that they can't get from textual material.

This online data transformation project isn't only an effort to digitize and make historical documents more easily accessible, but rather this initiative significantly improves and enhances the original printed Vernon city directory user experience. Wherever possible, spelling errors were corrected, duplicate entries were removed and most importantly, street address name changes and house number shifts have been meticulously researched, captured, highlighted and enhanced with modern equivalents. Therefore, when, for example the current location of a historical record is in the middle of an intersection, the project's goal is to keep that historical fact known and traceable. Whether demolished and replaced with a wider road, or whether the original building is still standing, the online database will provide users the tools to learn more about each address, its occupants and their journeys from 1900 to 1950 and eventually beyond.

The printed Vernon city directory is made up of several sections. It has a residential section that is listed in two ways – by resident's surname, as well as by street address, in alphabetical order. It has a business section that is listed by business type, and by business name, and corresponding advertisements are interspersed throughout all of the directory pages. ***Figure 1*** shows entries from the 1919 Vernon directories.



82 VERNON DIRECTORIES LIMITED

Dyck
'' Waldemar (Justina), engr Rubber Machy Shops, h 48 Theresa
Dyer, Margt, wks Mutual Life, 22 Heins
'' Mary (wid Robt), h 77 Strange
'' Robt W, wks Merch Rbr, 77 Strange
Dykeman, V Lorne (Ivy W), Chief Clerk C P Tickets and Communications, h 87 Krug, Phone 659M
Dziura, John (Vera), gro and serv stn 221 Ahrens w, h 91 Blucher
'' Mary, exam Wiegand & Co, 91 Blucher
'' Michl, wks Dumarts, 91 Blucher

E

E S Shoe Co Ltd, Helmuth Eisfeld pres; mfrs, 150 King e
Eadie, Elmer F, clk Bk of Montreal, 34 Pearl pl
Eagar, Cyril W (Edna M), wks Merch Rbr, 19 Vernon av
'' Elizbth (wid Robt J), h 56 Scott
'' Fredk, ctr Merch Rbr, 38 Strange
'' Walter J (Florence), btmkr Merch Rbr, h 38 Strange
'' Wilfd, wks Merch Rbr, 38 Strange
Eager, Ernest, lab, 26 Adam
Earl, David (Lillian), firemn Twin City Lndry, h 36 Fischer

Eaton
'' Andw J (A Irene), letter carrier, h 140 Peter
'' Byron W (S Elizbth), h 144 Homewood av
'' Clifford, drftsmn Dom Truck Equip 144 Homewood av
'' Eric S (Marie), supt Campbell Machy Co, h 153 Weber e
'' Sidney (Frances), mach Campbell Machy, h 21 Fairview av
'' T, Co Ltd, (Mail Order Dept), 57-63 King w
'' Violet, 21 Fairview av
Ebbage, W Harrison (Olive), opr News Record, h 455 Queen s
Ebel, Arnold, wks Beaver Furn, 36 Homewood av
'' Beatrice A, bkpr Merch Prntg, h 50 College, apt D
'' Jacob (Annie), floor fnshr, h 36 Homewood av
'' Stuart W (Feme I), wks McBrine Co, h 36 Homewood av
Eberhardt, John, wdwkr Dom Electrohome, 22 Courtland av e
Eberle, Carl, wks Schneiders, R R 3
'' Rose, wks Schneiders, R R 3
Eby, Aaron G (Hilda), wks Downing, Steen & Co, h 16 Brubacher
'' Ada B, tchr K & W Coll & Voc Schl, h 61 Benton, apt 2

Figure 1. Entries from 1919 Kitchener-Waterloo Vernon Directory

Process and Discovery

What appeared at first to be a simple project that listed people and businesses and their associated addresses, became significantly more complicated very quickly. Shifting city boundaries, street name and address number changes, Optical Character Recognition (OCR) errors, as well as data entry inaccuracies in the directories themselves needed to be studied and corrected. The project required a lot of manual input, continuous quality checks, external research, and the realization that the project team was discovering hurdles that no historical or technical resource could assist with.

Original Data Collection and Entry

Information available in the city directories is only as good as how it was collected and entered. Every year, every house was visited and surveyed on household members and their occupations and places of employment. Names were often recorded phonetically leading to a wide range of varying first and last name spelling throughout the years. For example, Henry Brose vs Hy Brox, or Snyder vs Schneider, or Bernofski vs Bernofocay. The latter truly demonstrating some of the problems with collecting information either orally, or from third parties. When data collectors were not able to speak to a member of the household, they would occasionally collect information from the neighbours. This would no doubt result in higher occurrences of misspellings, or extremely vague details, like "surname: Italians", or "the foreigners". Additionally, nicknames were also used, diluting the accuracy even further (i.e. William vs Willie). Spelling names incorrectly or inconsistently no doubt makes the online searching or look-up process more problematic and puts the onus on the researcher to try alternative spellings. ***Figure 2*** shows entries in the directory that highlight the fact that there may be alternative surname spellings.

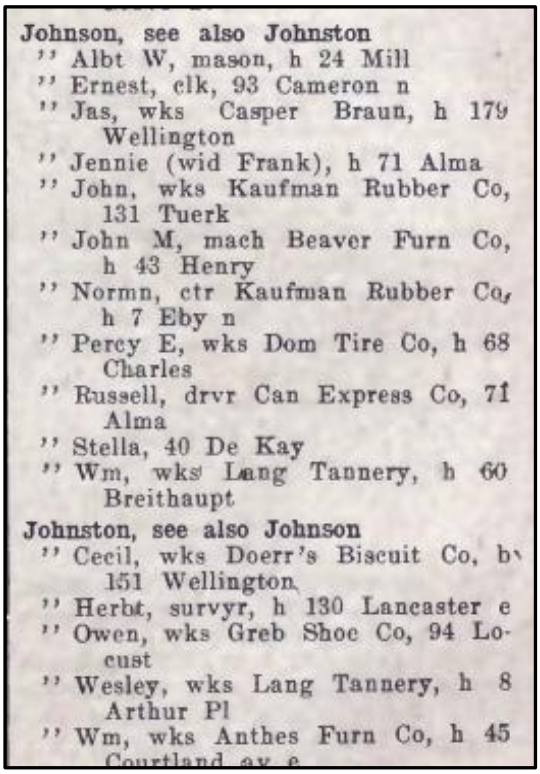

Johnson, see also Johnston
'' Albt W, mason, h 24 Mill
'' Ernest, clk, 93 Cameron n
'' Jas, wks Casper Braun, h 179 Wellington
'' Jennie (wid Frank), h 71 Alma
'' John, wks Kaufman Rubber Co, 131 Tuerk
'' John M, mach Beaver Furn Co, h 43 Henry
'' Normn, ctr Kaufman Rubber Co, h 7 Eby n
'' Percy E, wks Dom Tire Co, h 68 Charles
'' Russell, drvr Can Express Co, 71 Alma
'' Stella, 40 De Kay
'' Wm, wks Lang Tannery, h 60 Breithaupt
Johnston, see also Johnson
'' Cecil, wks Doerr's Biscuit Co, h 151 Wellington
'' Herbt, survyr, h 130 Lancaster e
'' Owen, wks Greb Shoe Co, 94 Locust
'' Wesley, wks Lang Tannery, h 8 Arthur Pl
'' Wm, wks Anthes Furn Co, h 45 Courtland av e

Figure 2. Example of different surname spellings in directory

The occupation and employment information was also at times quite vague. Sometimes all that could be obtained was "laborer", or "works at rubber company". Occupations weren't very consistent either. Sometimes "bookkeepers" were "ledgerkeepers". Many times abbreviations were used for occupations, so as a value-add and to assist with searching, project members expanded many abbreviations to full words (i.e. tchr = teacher, sec hd = second hand, mcht = machinist).

The discussion around street information accuracy is addressed in more detail below, however when it came to data entry, some years had the street name entered incorrectly. Project members corrected for this, however if street numbers were entered incorrectly, these likely were not caught and corrected since there is no way of knowing what was correct or not at the time of publication. Common errors were essentially just spelling mistakes – 104, instead of 164 or 184. If the geocoder matched the incorrect address, then it became part of the online database.

Data Conversion

When the original printed city directories were scanned using OCR software, it converted the printed characters into digital text. This allows users to search, copy, and paste the text into documents or spreadsheets. The OCR software, however, did not recognize all of the original characters, leading to several errors that resulted primarily in spelling mistakes of first and last names. Most errors were found in words that had "v" which translated to "y", and "e" which translated to "c". Just by scanning the names, it wasn't always possible to catch these types of errors. Occasionally, poor scanning also results in odd characters or missing letters. ***Figure 3*** shows an example of this.

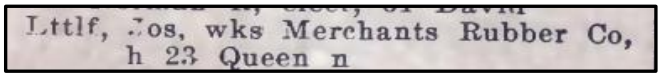


Figure 3. Poor scanning often leads to OCR conversion errors

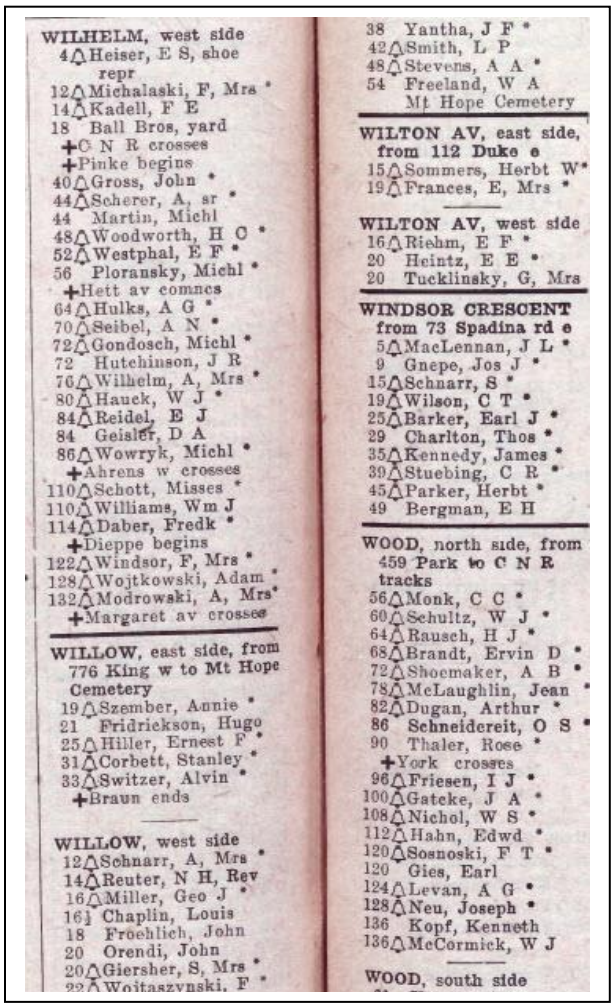
Data Inclusion

The purpose of this project was to map the information, and to do so required having an exact address of the household or business. If the city directory did not include an address, or if the address was vague, then that record had to be omitted. The early 1900s proved to be the most difficult in converting historical addresses to modern ones and efforts were taken to compare entries to subsequent years to infer correct addresses. Some examples of vague addresses included having only a street name without a house number, having only a description, "end of the street", being a Rural Route (i.e. RR 3) or having an address that was outside of the project's scope (i.e. in Waterloo, Conestogo, West Montrose, etc).

Street Matching

House numbers and street names have changed considerably over the years. What once was Dan Street, is now Wellington North, and Pinke is now Weber. In the early 1900s there was John Street, which is unrelated to the current John Street. Finding what was and what is now is certainly a game of connecting the dots and following all bread crumbs to clues about potential relations. Because of this, the project team developed two lists of streets – a current one and a historical one. The historical street addresses were created by analyzing all of the city directory street changes (usually by following a household over time), and confirming, when possible, using historical maps and fire insurance plans. Sometimes the directories highlighted street name changes with "Pinke is now Weber", but other times the street name just disappeared from the directory without notification. ***Figure 4*** shows an entry from the 1954 directory when Willow street still existed. ***Figure 5*** shows the 1955 directory where it was removed. After some significant time and research, project team members found a historical map (***Figure 6***) that shows the location of Willow street. Knowing this allowed the team to look at other maps to determine that Willow in

fact changed to Linwood street, which was eventually demolished and overbuilt and replaced with a parking lot (***Figure 7***).



WILHELM, west side
4 Heiser, E S, shoe repr
12 Michalaski, F, Mrs
14 Kadell, F E
18 Ball Bros, yard
+C N R crosses
+Pinke begins
40 Gross, John
44 Scherer, A, sr
44 Martin, Michl
48 Woodworth, H C
52 Westphal, E F
56 Ploransky, Michl
+Hett av comncs
64 Hulks, A G
70 Seibel, A N
72 Gondosch, Michl
72 Hutchinson, J R
76 Wilhelm, A, Mrs
80 Hauck, W J
84 Reidel, E J
84 Geisler, D A
86 Wowryk, Michl
+Ahrens w crosses
110 Schott, Misses
110 Williams, Wm J
114 Daber, Fredk
+Dieppe begins
122 Windsor, F, Mrs
128 Wojtkowski, Adam
132 Modrowski, A, Mrs
+Margaret av crosses

WILLOW, east side, from 776 King w to Mt Hope Cemetery
19 Szember, Annie
21 Fridrickson, Hugo
25 Hiller, Ernest F
31 Corbett, Stanley
33 Switzer, Alvin
+Braun ends

WILLOW, west side
12 Schnarr, A, Mrs
14 Reuter, N H, Rev
16 Miller, Geo J
16½ Chaplin, Louis
18 Froehlich, John
20 Orendi, John
20 Giersher, S, Mrs

38 Yantha, J F
42 Smith, L P
48 Stevens, A A
54 Freeland, W A
Mt Hope Cemetery

WILTON AV, east side, from 112 Duke e
15 Sommers, Herbt W
19 Frances, E, Mrs

WILTON AV, west side
16 Riehm, E F
20 Heintz, E E
20 Tucklinsky, G, Mrs

WINDSOR CRESCENT from 73 Spadina rd e
5 MacLennan, J L
9 Gnepe, Jos J
15 Schnarr, S
19 Wilson, C T
25 Barker, Earl J
29 Charlton, Thos
35 Kennedy, James
39 Stuebing, C R
45 Parker, Herbt
49 Bergman, E H

WOOD, north side, from 459 Park to C N R tracks
56 Monk, C C
60 Schultz, W J
64 Rausch, H J
68 Brandt, Ervin D
72 Shoemaker, A B
78 McLaughlin, Jean
82 Dugan, Arthur
86 Schneidereit, O S
90 Thaler, Rose
+York crosses
96 Friesen, I J
100 Gateke, J A
108 Nichol, W S
112 Hahn, Edwd
120 Sosnoski, F T
120 Gies, Earl
124 Levan, A G
128 Neu, Joseph
136 Kopf, Kenneth
136 McCormick, W J

WOOD, south side

Figure 4. In 1954 Willow street was listed

CITY OF KITCHENER 535

54 Fink, Geo H
66 Lindner, Wilfd
72 Gross, Cyril

WILHELM, east side, from 134 Edward to Margaret av
11 Kowalski, J C
11 Ausman, C
15 Roth, L, Mrs
15 Maranta, Gino
19 Pletch, S C
23 Krawczyk, M, Mrs
23 Kotecki, John
27 Krawczyk, Louis,
27 Mehlgarten, John
31 Psutka, Joe F
+C N R crosses
39 Zablocki, John
43 Sandrin, Angelo
51 Markowsky, R
57 Forler, A P
57 Lerch, Florence
57 Carter, Dawson
63 Profendiver, Ernst
63 Wolfe, Gordon
63 Riedl, Jos, tnsmth
+Augusta begins
67 Helwig, W G
69 Boegel, I, Mrs
75 Stuckhardt, Calvin
75 Michalski, John
83 Lamb, A G
85 Magier, Florian
91 Graf, Anthony J
95 Rieck, Wm H
+Ahrens w crosses
109 Erdman, Anatol
109 Vacant apt
113 Krich, Matthias
113 Dienesck, Michl
113 Rochan, Herman
115 Baird, Harry
121 Tucker, Percy
125 Tosheff, Nichls
129 Richter, A

14 Vacant
18 Ball Bros, yard
+C N R crosses
+Pinke begins
40 Gross, John
44 Scherer, A, sr
44 Martin, Michl
48 Woodworth, H C
52 Westphal, E F
56 Ploransky, Michl
+Hett av comncs
64 Hulks, A G
70 Seibel, A N
72 Gondosch, Michl
72 Hutchinson, J R
76 Wilhelm, A, Mrs
80 Hauck, W J
84 Reidel, E J
84 Vacant apt
86 Wowryk, Michl
+Ahrens w crosses
110 Schott, Misses
110 Williams, Wm J
114 Daber, Fredk
+Dieppe begins
122 Windsor, F, Mrs
128 Wojtkowski, Adam
132 Modrowski, A, Mrs
+Margaret av crosses

WILSON, from Mill
New houses (4)

WILTON AV, east side, from 112 Duke e
15 Sommers, Herbt W
19 Reiter, Michl

WILTON AV, west side
16 Riehm, E F
16 Weber, W G
20 Heintz, E E
20 Tucklinsky, G, Mrs

WINDSOR CRESCENT from 73 Spadina rd e

Figure 5. In 1955 Willow street was not listed

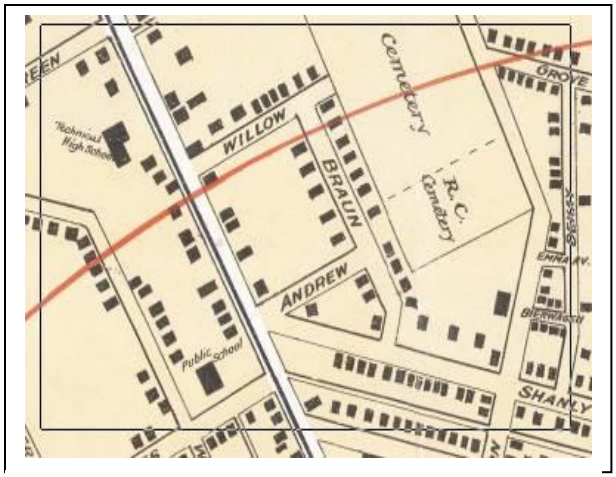


Figure 6. A map from 1913 showing Willow Street

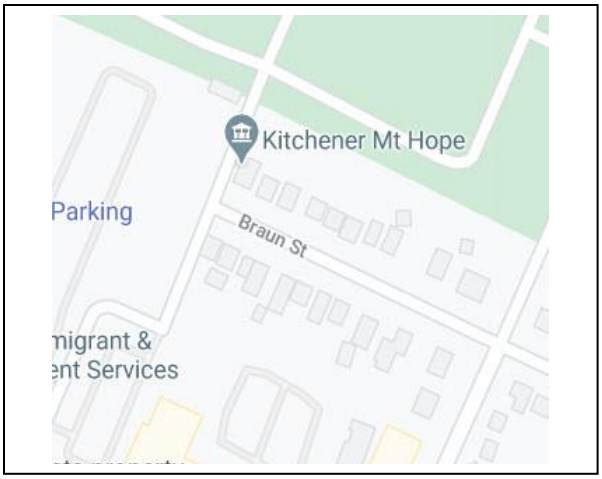
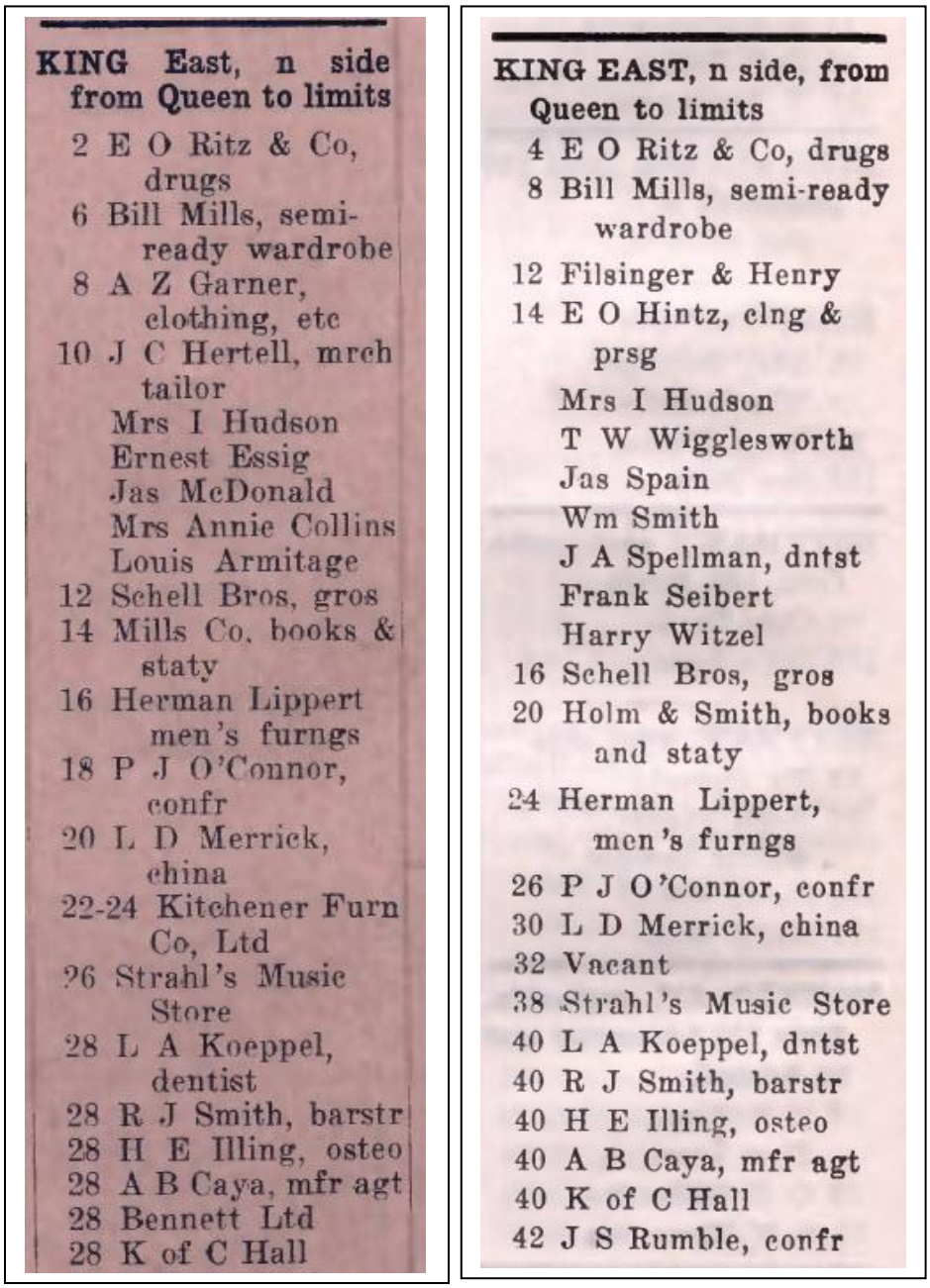


Figure 7. A map from 2021 showing a parking lot

These changes and non-standard address formats were the main drivers of errors in this project since some street names changed several times over 120 years with some having the same names of streets that didn't change. In several instances when the street name changed, the address numbers changed as well (i.e. 3 Pinke = 5 Weber). To muddy the waters even further, on several occasions street numbers changed without the street name changing. ***Figure 8*** shows how King Street East shifted between 1922 and 1923. These findings were often discovered serendipitously, enforcing the requirement to meticulously follow each household over the years to create a matching 'address shift lookup table'.



KING East, n side from Queen to limits

2 E O Ritz & Co, drugs
6 Bill Mills, semi-ready wardrobe
8 A Z Garner, clothing, etc
10 J C Hertell, mrch tailor
Mrs I Hudson
Ernest Essig
Jas McDonald
Mrs Annie Collins
Louis Armitage
12 Schell Bros, gros
14 Mills Co, books & staty
16 Herman Lippert men's furngs
18 P J O'Connor, confr
20 L D Merrick, china
22-24 Kitchener Furn Co, Ltd
26 Strahl's Music Store
28 L A Koeppel, dentist
28 R J Smith, barstr
28 H E Illing, osteo
28 A B Caya, mfr agt
28 Bennett Ltd
28 K of C Hall

KING EAST, n side, from Queen to limits

4 E O Ritz & Co, drugs
8 Bill Mills, semi-ready wardrobe
12 Filsinger & Henry
14 E O Hintz, clng & prsg
Mrs I Hudson
T W Wigglesworth
Jas Spain
Wm Smith
J A Spellman, dntst
Frank Seibert
Harry Witzel
16 Schell Bros, gros
20 Holm & Smith, books and staty
24 Herman Lippert, men's furngs
26 P J O'Connor, confr
30 L D Merrick, china
32 Vacant
38 Strahl's Music Store
40 L A Koeppel, dntst
40 R J Smith, barstr
40 H E Illing, osteo
40 A B Caya, mfr agt
40 K of C Hall
42 J S Rumble, confr

Figure 8. 1922 (left) and 1923 (right). King Street East address numbers had shifted

It should also be mentioned that cardinal changes do affect modern street addresses. What may at one point have been Victoria Street, today is Victoria Street North, and as such needed to be updated so records could be placed correctly on the map. Due to the thousands of records affected by these changes, there is no guarantee that they were all caught. When team members reached out to the city records office to inquire about a list of address changes, they did not have one that included address changes before the 1950s. The Geospatial Centre plans to compile the list and share it with the city along with others who may be interested.

Changing Landscape

It goes without saying that Berlin in the early 1900s looked much different than Kitchener does today. Every year, fewer original buildings stand as they continue to be replaced by transportation infrastructure, high rises and business districts. Because of this, very few original addresses from the turn of the 20th century can be accurately mapped. Instead of omitting these from the database, project team members did their best to add them in using modern day's location. Using the previous example of Willow street being replaced by a parking lot, staff have recorded the approximate geographic coordinate of each home. A modern address doesn't actually exist, so the user would only be able to search by the historical address (i.e. Willow). These records actually have an additional field added, OANF (Original Address Not Found), flagging these as not having a true spot on the modern map as we know it, but yet, they are not forgotten. This additional field will also help researchers find all homes that were demolished and not replaced by the same street number.

Shifting City Boundaries

The scope of this project is on Kitchener, even though the city directories do also include Waterloo. Historical Waterloo actually includes some sections of modern day Kitchener. Likewise, what was considered to be Kitchener 100 years ago (i.e. King St N), is modern day Waterloo. The geocoding process filtered out modern day addresses that are in Waterloo, however for those addresses that were not originally part of Kitchener, but were in the Waterloo section of the directory, manual additions were required.

Methodology

This section will summarize some of the more relevant aspects of the team's technical methodology, addressing the many issues discussed in the previous section.

Data Transformation and Cleanup

As briefly discussed, the process of scanning documents using OCR picks up many artifacts that needed to be filtered out. This was especially present in the 1900-1930 volumes of the directory. After the initial discovery and inventory of character discrepancies, staff were provided with a list to look out for when transcribing the data from pdf to csv. ***Figure 9*** is an example of some of the characters that needed to be converted.

OCR - problems

*Uppercase **W's** as AA7, AV, A\, AAr*

*Uppercase **M's** as A\, Al e.g. Alill = **Mill** or Alii ler =**Miller***

*Lowercase **h** as li e.g. Wliitewear = **Whitewear** or merli = **merch***

*Uppercase **Y's** as \r*

*Lowercase **y** as v, ^*

*Uppercase **V** as *

*Lowercase **w** as Av*

*Lowercase **u** as n*

***Du** as Dn*

***Mn** as Mu*

***1)** as D*

***H** as Il or IT or II*

***X** as N*

***'b** as 's*

Figure 9. Documentation shared with staff transcribing pdf to csv

There were many attempts at automation using Python, VBScript, Excel formulas and NotePad ++ text editing tools. Unfortunately, there were so many variables that the time it was taking to code something to correct the OCR artifacts was becoming longer than transferring the text line by line. ***Figure 10*** shows sample Python code, ***Figure 11*** shows a sample of VBScript code, ***Figure 12*** shows an Excel sample, and finally ***Figure 13*** shows a NotePad ++ sample.

```
# If any LAParams group arguments were passed, create an LAParams object and
    # populate with given args. Otherwise, set it to None.
    if not no_laparams:
        laparams = pdfminer.layout.LAParams()
        for param in ("all_texts", "detect_vertical", "word_margin", "char_margin", "line_margin", "boxes_flow"):
            paramv = locals().get(param, None)
            if paramv is not None:
                setattr(laparams, param, paramv)
    else:
        laparams = None

    imagewriter = None
    if output_dir:
        imagewriter = ImageWriter(output_dir)

    if output_type == "text" and outfile != "-":
        for override, alttype in (  (".htm", "html"),
                                    (".html", "html"),
                                    (".xml", "xml"),
                                    (".tag", "tag") ):
            if outfile.endswith(override):
                output_type = alttype

    if outfile == "-":
        outfp = sys.stdout
        if outfp.encoding is not None:
            codec = 'utf-8'
    else:
        outfp = open(outfile, "wb")

    for fname in files:
        with open(fname, "rb") as fp:
            pdfminer.high_level.extract_text_to_fp(fp, **locals())
    return outfp

def maketheparser():
    parser = argparse.ArgumentParser(description=__doc__, add_help=True)
    parser.add_argument("files", type=str, default=None, nargs="+", help="File to process.")
    parser.add_argument("-d", "--debug", default=False, action="store_true", help="Debug output.")
    parser.add_argument("-p", "--pagenos", type=str, help="Comma-separated list of page numbers to parse. Included for legacy
applications, use --page-numbers for more idiomatic argument entry.")
    parser.add_argument("--page-numbers", type=int, default=None, nargs="+", help="Alternative to --pagenos with space-
separated numbers; supercedes --pagenos where it is used.")
    parser.add_argument("-m", "--maxpages", type=int, default=0, help="Maximum pages to parse")
    parser.add_argument("-P", "--password", type=str, default="", help="Decryption password for PDF")
    parser.add_argument("-o", "--outfile", type=str, default="-", help="Output file (default \"-\" is stdout)")
    parser.add_argument("-t", "--output_type", type=str, default="text", help="Output type: text|html|xml|tag (default is text)")
    parser.add_argument("-c", "--codec", type=str, default="utf-8", help="Text encoding")
    parser.add_argument("-s", "--scale", type=float, default=1.0, help="Scale")
    parser.add_argument("-A", "--all-texts", default=None, action="store_true", help="LAParams all texts")
    parser.add_argument("-V", "--detect-vertical", default=None, action="store_true", help="LAParams detect vertical")
    parser.add_argument("-W", "--word-margin", type=float, default=None, help="LAParams word margin")
    parser.add_argument("-M", "--char-margin", type=float, default=None, help="LAParams char margin")
    parser.add_argument("-L", "--line-margin", type=float, default=None, help="LAParams line margin")
    parser.add_argument("-F", "--boxes-flow", type=float, default=None, help="LAParams boxes flow")
```

Figure 10. This code read from the pdf's and placed the "cleaned" text into a csv file through Excel

```
'find last row
lRow = Cells(Rows.Count, 1).End(xlUp).Row

'rCells = Worksheets("test").Cells

'We set our range variable = cell A1 in the active sheet.
Set rCell = Range("A1")
For i = 3 To lRow
   If InStr(Cells(i, 1).Value, ",") = 0 Then
      Cells(i - 1, 1).Value = Cells(i - 1, 1).Value & " " & Cells(i, 1).Value
      Cells(i, 1).Value = ""
   End If
Next i
For i = 3 To lRow
   If Cells(i, 1).Value = "" Then
      Cells(i, 1).Value = Cells(i + 1, 1).Value
      Cells(i + 1, 1).Value = Cells(i + 2, 1).Value
   End If
Next i
For i = 3 To lRow
   If Cells(i, 1).Value = "" Then
      Cells(i, 1).Value = Cells(i + 1, 1).Value
      Cells(i + 1, 1).Value = Cells(i + 2, 1).Value
   End If
Next i
For i = 3 To lRow
   If Cells(i, 1).Value = "" Then
      Cells(i, 1).Value = Cells(i + 1, 1).Value
      Cells(i + 1, 1).Value = Cells(i + 2, 1).Value
   End If
Next i
For i = 3 To lRow
   If Cells(i, 1).Value = "" Then
      Cells(i, 1).Value = Cells(i + 1, 1).Value
      Cells(i + 1, 1).Value = Cells(i + 2, 1).Value
   End If
Next i
For i = 3 To lRow
   If Cells(i, 1).Value = "" Then
      Cells(i, 1).Value = Cells(i + 1, 1).Value
      Cells(i + 1, 1).Value = Cells(i + 2, 1).Value
   End If
Next i
For i = 2 To lRow
   If (InStr(Cells(i, 1).Value, "Ltd,") > 0 And InStr(Cells(i, 1).Value, "Ltd,") < 10) Or (InStr(Cells(i, 1).Value, "Co,") > 0 And InStr(Cells(i, 1).Value,
"Co,") < 10) Then
      Cells(i - 1, 1).Value = Cells(i - 1, 1).Value & " " & Cells(i, 1).Value
      Cells(i, 1).Value = ""
   End If
Next i

'
' -------------------------------step2 Macro---------------------------------
'
```

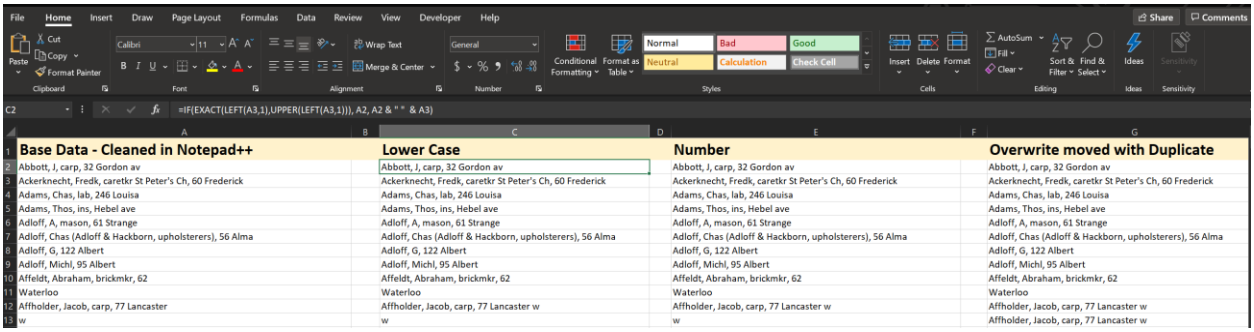
Figure 11. This code read from a copy and dumped into a text file and imported that into Excel

Excel Sample:

=IF(EXACT(LEFT(A3,1),UPPER(LEFT(A3,1))), A2, A2 & " " & A3)

=IF(ISERR(LEFT(C3,1)*1),C2,C2 & " " & C3)

=IF(COUNT(FIND({0,1,2,3,4,5,6,7,8,9},A18))>0,,A18)



Base Data - Cleaned in Notepad++	Lower Case	Number	Overwrite moved with Duplicate
Abbott, J, carp, 32 Gordon av	Abbott, J, carp, 32 Gordon av	Abbott, J, carp, 32 Gordon av	Abbott, J, carp, 32 Gordon av
Ackerknecht, Fredk, caretkr St Peter's Ch, 60 Frederick	Ackerknecht, Fredk, caretkr St Peter's Ch, 60 Frederick	Ackerknecht, Fredk, caretkr St Peter's Ch, 60 Frederick	Ackerknecht, Fredk, caretkr St Peter's Ch, 60 Frederick
Adams, Chas, lab, 246 Louisa	Adams, Chas, lab, 246 Louisa	Adams, Chas, lab, 246 Louisa	Adams, Chas, lab, 246 Louisa
Adams, Thos, ins, Hebel ave	Adams, Thos, ins, Hebel ave	Adams, Thos, ins, Hebel ave	Adams, Thos, ins, Hebel ave
Adloff, A, mason, 61 Strange	Adloff, A, mason, 61 Strange	Adloff, A, mason, 61 Strange	Adloff, A, mason, 61 Strange
Adloff, Chas (Adloff & Hackborn, upholsterers), 56 Alma	Adloff, Chas (Adloff & Hackborn, upholsterers), 56 Alma	Adloff, Chas (Adloff & Hackborn, upholsterers), 56 Alma	Adloff, Chas (Adloff & Hackborn, upholsterers), 56 Alma
Adloff, G, 122 Albert	Adloff, G, 122 Albert	Adloff, G, 122 Albert	Adloff, G, 122 Albert
Adloff, Michl, 95 Albert	Adloff, Michl, 95 Albert	Adloff, Michl, 95 Albert	Adloff, Michl, 95 Albert
Affeldt, Abraham, brickmkr, 62	Affeldt, Abraham, brickmkr, 62	Affeldt, Abraham, brickmkr, 62	Affeldt, Abraham, brickmkr, 62
Waterloo	Waterloo	Waterloo	Waterloo
Affholder, Jacob, carp, 77 Lancaster	Affholder, Jacob, carp, 77 Lancaster w	Affholder, Jacob, carp, 77 Lancaster w	Affholder, Jacob, carp, 77 Lancaster w
w	w	w	Affholder, Jacob, carp, 77 Lancaster w

Figure 12. Excel sample. This was an attempt to dump the raw OCR'd text and clean it column by column using formulas

How to handle names w/o numbers on streets?
Add numbers starting with 1 (short new streets) if it makes sense ... first check if you might be in the middle of a street.

1.Go to the search menu Ctrl + F and open the "Mark" tab.
2.Check "Bookmark line" (if there is no "Mark" tab update to the current version).
3.Enter the +
4.Now go to the Menu "Search -> Bookmark -> Remove Bookmarked lines"
5.Done.

The "r" number problem
Find, Mark,Remove bookmarks
^[a-q,s-z]

Also, To find a letter followed by a number
[a-z][0-9]+

Find 'not' numbers first caracter... duplicate street names too
Find, Mark,Check 'n Clean..... Remove bookmarks
^[A-Z]

Figure 13. This was a mass copy from the pdf's into NotePad ++ and using the text editor tools within to clean the OCR'd text as much as possible before inputting into an Excel csv

All these automation tests were adequate for the specific years they were built in but since there was so much variation in artifacts and in layout among the years these efforts proved very difficult to take advantage of.

The team settled on transposing the directories into csv's line by line. This was straight forward and easy to explain to any staff member willing to work on the project.

Clean up

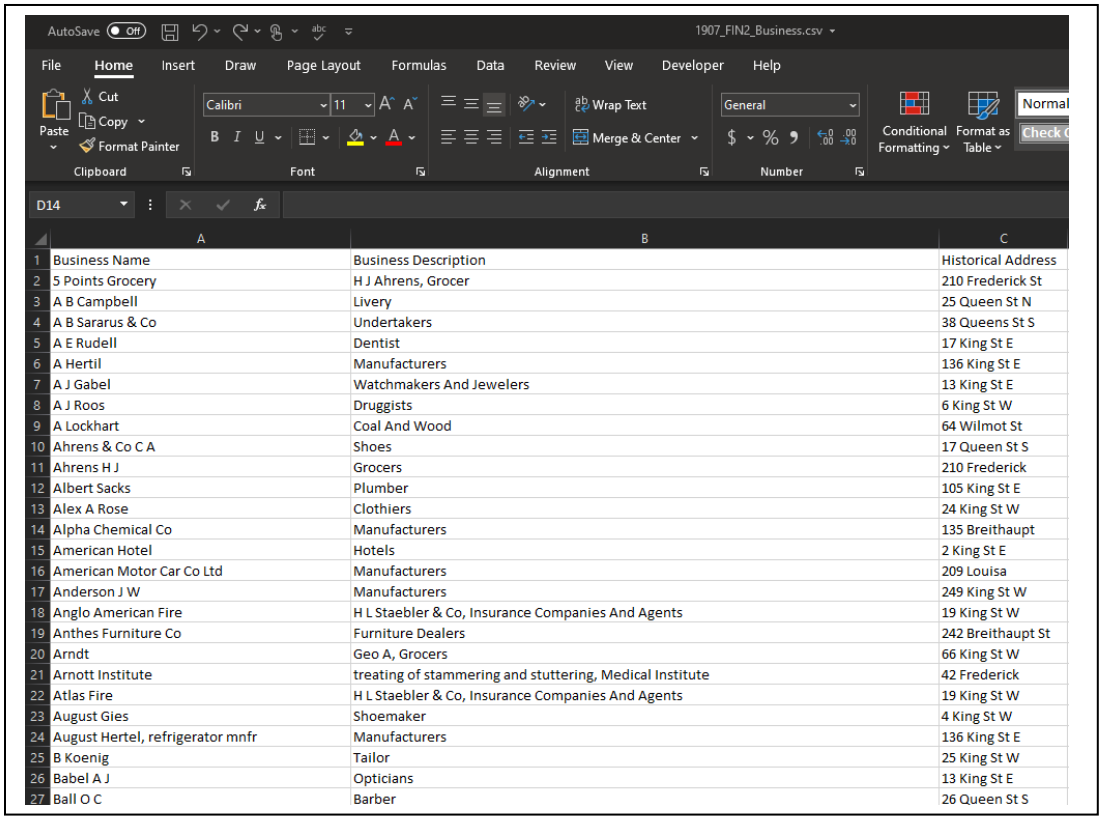
QA stage 1:

Go through all completed years manually and ensure that the text data is in the correct column.

Make sure:

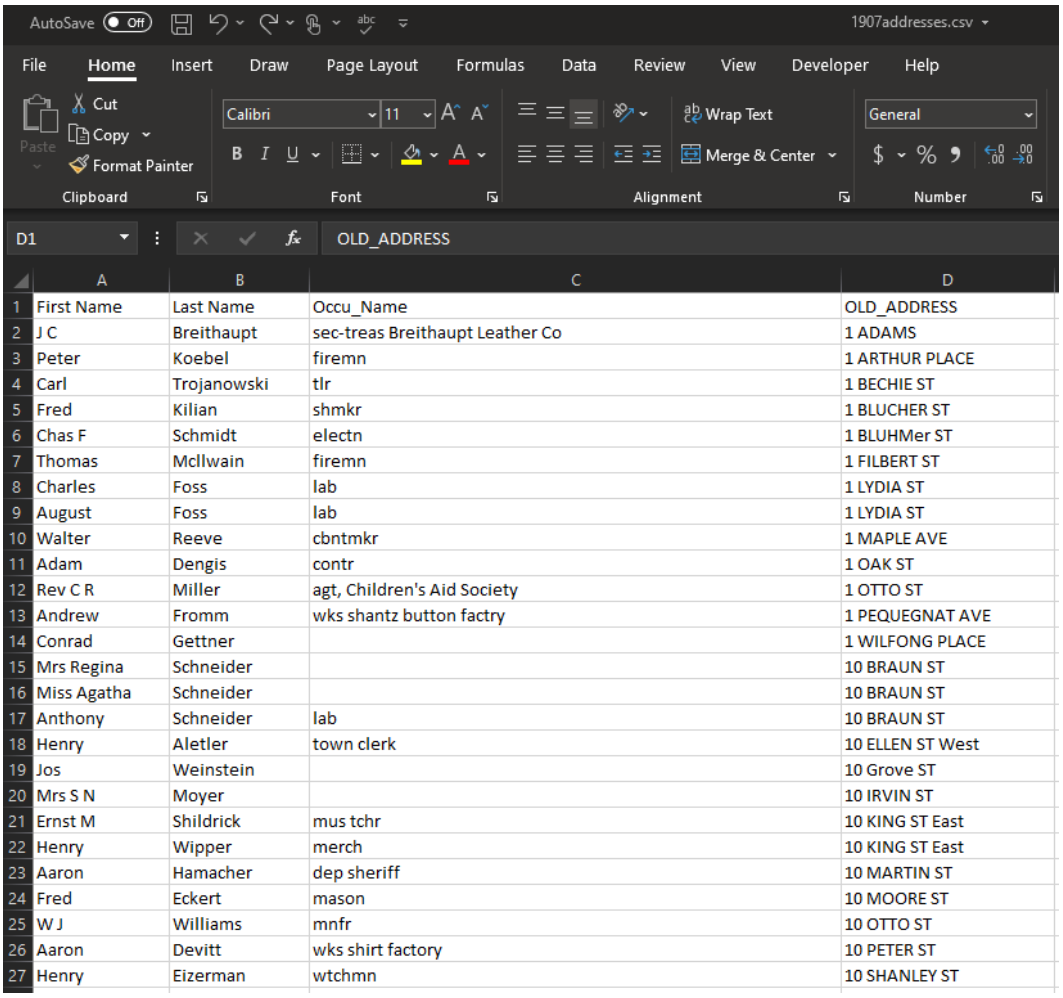
- The data is correct for each row
- There aren't businesses in the residential dataset
- The address format is correct for the later python Updater code (Number, Street, Abv, Cardinal)
- There are no empty rows, missing names, or addresses
- Remove Waterloo addresses
- General spelling and OCR artifact letter correction missed in the transposition
- Split the dataset into Business and People

Figure 14 shows an example of the business csv, and ***Figure 15*** shows the residential csv.



	A	B	C
1	Business Name	Business Description	Historical Address
2	5 Points Grocery	H J Ahrens, Grocer	210 Frederick St
3	A B Campbell	Livery	25 Queen St N
4	A B Sararus & Co	Undertakers	38 Queens St S
5	A E Rudell	Dentist	17 King St E
6	A Hertil	Manufacturers	136 King St E
7	A J Gabel	Watchmakers And Jewelers	13 King St E
8	A J Roos	Druggists	6 King St W
9	A Lockhart	Coal And Wood	64 Wilmot St
10	Ahrens & Co C A	Shoes	17 Queen St S
11	Ahrens H J	Grocers	210 Frederick
12	Albert Sacks	Plumber	105 King St E
13	Alex A Rose	Clothiers	24 King St W
14	Alpha Chemical Co	Manufacturers	135 Breithaupt
15	American Hotel	Hotels	2 King St E
16	American Motor Car Co Ltd	Manufacturers	209 Louisa
17	Anderson J W	Manufacturers	249 King St W
18	Anglo American Fire	H L Staebler & Co, Insurance Companies And Agents	19 King St W
19	Anthes Furniture Co	Furniture Dealers	242 Breithaupt St
20	Arndt	Geo A, Grocers	66 King St W
21	Arnott Institute	treating of stammering and stuttering, Medical Institute	42 Frederick
22	Atlas Fire	H L Staebler & Co, Insurance Companies And Agents	19 King St W
23	August Gies	Shoemaker	4 King St W
24	August Hertel, refrigerator mnfr	Manufacturers	136 King St E
25	B Koenig	Tailor	25 King St W
26	Babel A J	Opticians	13 King St E
27	Ball O C	Barber	26 Queen St S

Figure 14. Business csv showing the first three columns



	A	B	C	D
1	First Name	Last Name	Occu_Name	OLD_ADDRESS
2	J C	Breithaupt	sec-treas Breithaupt Leather Co	1 ADAMS
3	Peter	Koebel	firemn	1 ARTHUR PLACE
4	Carl	Trojanowski	tlr	1 BECHIE ST
5	Fred	Kilian	shmkr	1 BLUCHER ST
6	Chas F	Schmidt	electn	1 BLUHMer ST
7	Thomas	McIlwain	firemn	1 FILBERT ST
8	Charles	Foss	lab	1 LYDIA ST
9	August	Foss	lab	1 LYDIA ST
10	Walter	Reeve	cbntmkr	1 MAPLE AVE
11	Adam	Dengis	contr	1 OAK ST
12	Rev C R	Miller	agt, Children's Aid Society	1 OTTO ST
13	Andrew	Fromm	wks shantz button factry	1 PEQUEGNAT AVE
14	Conrad	Gettner		1 WILFONG PLACE
15	Mrs Regina	Schneider		10 BRAUN ST
16	Miss Agatha	Schneider		10 BRAUN ST
17	Anthony	Schneider	lab	10 BRAUN ST
18	Henry	Aletler	town clerk	10 ELLEN ST West
19	Jos	Weinstein		10 Grove ST
20	Mrs S N	Moyer		10 IRVIN ST
21	Ernst M	Shildrick	mus tchr	10 KING ST East
22	Henry	Wipper	merch	10 KING ST East
23	Aaron	Hamacher	dep sheriff	10 MARTIN ST
24	Fred	Eckert	mason	10 MOORE ST
25	W J	Williams	mnfr	10 OTTO ST
26	Aaron	Devitt	wks shirt factory	10 PETER ST
27	Henry	Eizerman	wtchmn	10 SHANLEY ST

Figure 15. Residential csv showing the first four columns

QA Stage 2 – Levenshtein Distance Stage 1:

Run the cleaned csv's through a python script that checks abbreviations of certain words and updates them to full words. Eg. Saw to Sawyer or Sec to Secretary. This is an effort to update old language to a modern equivalent and/or change the abbreviation to the full word for easier reading.

The Python script created is the first of two uses of the Levenshtein distance algorithm.

"The Levenshtein distance (a.k.a edit distance) is a measure of similarity between two strings. It is defined as the minimum number of changes required to convert string a into string b (this is done by inserting, deleting, or replacing a character in string a). The smaller the Levenshtein distance, the more similar the strings are" (Educative, 2021).

In this case the Levenshtein distance was used to read a column in the residential csv and compare the selected abbreviated words to the list of modern equivalents (***Figure 16***).

csv from Stage 1 QA:

sec-treas Breithaupt Leather Co
firemn
tlr
shmkr
electn
firemn
lab
lab
cbntmkr
contr
agt, Children's Aid Society
wks shantz button factry
lab
town clerk
mus tchr
merch
dep sheriff
mason
mnfr
wks shirt factory
wtchmn

Python code snippet:

```
# Replacing all split words that match with abbreviations
Employments = df_fw['Employment'].tolist()


edited_row = []
edited_col = []
var = 0

for rows in Employments:
    row_list = str(rows).split()

    for word in row_list:
        if re.search("[\w]+(?=\,)",word):
            word = re.sub("[,]", "", word)
            var = 1
            print(word)

        if word.strip().upper() in list_of_abbreviations:

            if var == 1:
                edited_row = edited_row + [word.replace(w
                var = 0

            else:
                edited_row = edited_row + [word.replace(
        else:
            if var == 1:
                edited_row = edited_row + [word + ',']
                var = 0
            else:
                edited_row = edited_row + [word]

    per_row = ' '.join(edited_row)
    edited_col = edited_col + [per_row]
    edited_row = []

df_fw['edited_Employments'] = edited_col
```

csv updated abv.:

secretary treasurer Breithaupt Leather Co
fireman
tailor
shoemaker
electn
fireman
laborer
laborer
cabinetmaker
contractor
agent, Children's Aid Society
works shantz button factry
laborer
town clerk
music teacher
merchant
deputy sheriff
masoner
manufacturer
works shirt factory
watchman

Figure 16. Using the Levenshtein distance for abbreviated words

Figure 17 shows the list of comparators the Levenshtein distance algorithm uses to find and change the abbreviations:

	A	B
1	AN	FN
2	acct	accountant
3	adjstr	adjuster
4	adjt	adjutant
5	admin	administrator
6	adv	advertising
7	agcy	agency
8	agt	agent
9	alter	alterations
10	Amer	American
11	Angln	Anglican
12	anncr	announcer
13	aplncs	appliances
14	appr	apprentice
15	apt	apartment
16	arch	architect
17	asmblr	assembler
18	assest	assessment
19	assn	association
20	assoc	associate
21	assr	assessor
22	asst	assistant

Figure 17. List of comparators

QA Stage 3:

Repeat QA Stage 1 with particular attention to the Historical Address column.

Add any missed abbreviations to the list of comparators that are deemed relevant for future years.

Re-run Levenshtein Distance Stage 1 with the updated list of comparators.

QA Stage 4 - Levenshtein Distance Stage 2:

With the abbreviations updated staff were able to focus on updating the addresses to modern ones. As discussed earlier, cities change quite a bit over the span of a hundred years, so the addresses needed to be modified accordingly so that they can be geolocated in the correct position on the map. This process is a little bit more involved than the *QA Stage 2* because the address dataset needs to be broken up into its constituent components (Number, Street Name, Street abbreviation, and cardinal direction). The Levenshtein Distance algorithm in the python code is only used on the street name part of the string. The rest of the string must be parsed and reassembled with either the same street name (if not changed) or the updated street name (if it is changed). Along with the Levenshtein Distance algorithm there are many checks and changes in this code to assemble the correct address for the geocoder. This Python code reads from the transposed csv and an Updater csv. The transposed csv gives the historical address and the Updater csv gives the Levenshtein Distance comparator addresses along with address not found flags and x and y coordinates. These flags and x and y are for addresses that don't have modern equivalents and have coordinates for where the buildings used to be. This is in an effort to minimize the repeat of mismatches in the geolocator for future years. Having the x and y coordinates or nearest modern addresses greatly reduces the rework needed to ensure all the historical address points show up correctly on a modern map. ***Figure 18*** shows a sample of the address checkers Levenshtein code.

```
# Calculating levenshtein ratio (distance between mispelled words) using a levenshtein ratio

def levenshtein_ratio(s, t, ratio_calc = True):
    """ levenshtein_ratio_and_distance:
        Calculates levenshtein distance between two strings.
        If ratio_calc = True, the function computes the
        levenshtein distance ratio of similarity between two strings
        For all i and j, distance[i,j] will contain the Levenshtein
        distance between the first i characters of s and the
        first j characters of t
    """
    # Initialize matrix of zeros
    rows = len(s)+1
    cols = len(t)+1
    distance = np.zeros((rows,cols),dtype = int)

    # Populate matrix of zeros with the indeces of each character of both strings
    for i in range(1, rows):
        for k in range(1,cols):
            distance[i][0] = i
            distance[0][k] = k

    # Iterate over the matrix to compute the cost of deletions,insertions and/or substitutions
    for col in range(1, cols):
        for row in range(1, rows):
            if s[row-1] == t[col-1]:
                cost = 0 # If the characters are the same in the two strings in a given position [i,j] then the cost is 0
            else:
                # In order to align the results with those of the Python Levenshtein package, if we choose to calculate the r
                # the cost of a substitution is 2. If we calculate just distance, then the cost of a substitution is 1.
                if ratio_calc == True:
                    cost = 2
                else:
                    cost = 1
            distance[row][col] = min(distance[row-1][col] + 1,      # Cost of deletions
                                     distance[row][col-1] + 1,      # Cost of insertions
                                     distance[row-1][col-1] + cost)     # Cost of substitutions
    if ratio_calc == True:
        # Computation of the Levenshtein Distance Ratio
        Ratio = ((len(s)+len(t)) - distance[row][col]) / (len(s)+len(t))
        return Ratio
    else:
        # print(distance) # Uncomment if you want to see the matrix showing how the algorithm computes the cost of deletions,
        # insertions and/or substitutions
        # This is the minimum number of edits needed to convert string a to string b
        return distance[row][col]
```

Figure 18. Sample of the address checkers Levenshtein code

Since there are going to be at least a million rows of data in this directory database and they all must be geocoded using the Levenstein ratio, this gives the best chance to find, replace and correct any addresses that may be misspelled or similar. Street names with the same first characters, like Albert and Alberta, must be accounted for, as they are two distinct names but in a normal string comparison search may be perceived as the same. This is exemplified in searches done on any number of webpages where one might be using a search term that is short and to the point, but the returns are long and multi-word incorporating the input term. Text autocorrect may also be an example when one doesn't exactly know the spelling and gets back only the words made up of the typed letters because the algorithm only compares literal strings to its database of words.

In ***Figure 19*** the historical address (1908) was passed through the Levenstein Python code, parsed and compared to the Updater csv (***Figure 20***) resulting in the Current Address column being generated.

	A	B	C	D
1	Business Type	Business Name	Historical Address	Current Address
2	Association	Young Womens Christian Ass	72 Queen S	72 Queen St S
3	Music Store	Wanless G A	20 King W	20 King St W
4	Music	W A Phillmore Music	76 King St E	150 King St E
5	PIANOS AND ORGANS	Gardiner T G	29 Queen S	29 Queen St S
6	PIANOS AND ORGANS	Bressnahan J P	67 King E	125 King St E
7	PIANOS AND ORGANS	Gourlay Piano	84 King W	84 King St W
8	Shirt Manufacturer	Harry Tollon	80 King St E	154 King St E
9	Accountants	Scully & Scully	9 Foundry S	9 Foundry St S
10	Grocer	John Waldschmidt	82 King St W	82 King St W
11	ARCHITECTS	Wey I	21 King E	33 King St E
12	ARCHITECTS	Knechtel Chas	31 Courtland W	31 Courtland W
13	AUCTIONEER	Dey J W	59 King E	117 King St E
14	BAKERS	Ferguson Fred K	35 Queen N	35 Queen St N
15	BAKERS	Dey John W	59 King E	117 King St E
16	BAKERS	Bardon Louis	64 Foundry S	120 Ontario St S
17	BAKERS	Dietrich H A	86 King E	86 King St E
18	Bank	Bank of Hamilton	1 King E	1 King St E
19	BANKS	Bank of Toronto	1 King W	1 King St E
20	BANKS	Merchants Bank	46 King W	46 King St W
21	BANKS	Bank of Commerce	48 King E	70 King St E
22	BANKS	Bank of Nova Scotia	72 King W	72 King St W
23	Banks	Dominion Bank	45 King E	45 King St E
24	BARRISTER and SOLICITORS	Cram W M	13 Weber E	13 Weber St E
25	BARRISTER and SOLICITORS	Bowlby D S	16 Margaret	16 Queen St N
26	BARRISTER and SOLICITORS	Reade W M	18 Queen N	18 Queen St N
27	BARRISTER and SOLICITORS	Bowlby W H	221 King W	221 King St E

Figure 19. 1908 csv with Current Address automatically added

	A	B	C	D	E	F	G
1	Original St Name	Original St Number	Fromer Complete Address	Modern St Name	Modern St Number	Result	Shortened Original St Name
2	Albert St N	25	25 Albert St N	Madison Ave N	25	25 Madison Ave N	ALBERT N
3	Albert St N	45	45 Albert St N	Madison Ave N	45	45 Madison Ave N	ALBERT N
4	Albert St N	49	49 Albert St N	Madison Ave N	49	49 Madison Ave N	ALBERT N
5	Albert St N	53	53 Albert St N	Madison Ave N	53	53 Madison Ave N	ALBERT N
6	Albert St N	57	57 Albert St N	Madison Ave N	57	57 Madison Ave N	ALBERT N
7	Albert St N	61	61 Albert St N	Madison Ave N	61	61 Madison Ave N	ALBERT N
8	Albert St N	65	65 Albert St N	Madison Ave N	65	65 Madison Ave N	ALBERT N
9	Albert St N	10	10 Albert St N	Madison Ave N	10	10 Madison Ave N	ALBERT N
10	Albert St N	14	14 Albert St N	Madison Ave N	14	14 Madison Ave N	ALBERT N
11	Albert St N	18	18 Albert St N	Madison Ave N	18	18 Madison Ave N	ALBERT N
12	Albert St N	22	22 Albert St N	Madison Ave N	22	22 Madison Ave N	ALBERT N

Figure 20. Updater csv

Within the Address Updater Python code is also a check for addresses that do not have a modern equivalent or the location of the nearest modern address is far from where the historical address was. This check adds a flag (OANF – Original Address Not Found) in a Note column (***Figure 21***). This flag is a statement that the modern geocodable address is close to the original for some cases or if it is accompanied by X and Y coordinates the address was created by staff to accommodate the lack of a modern location. These flags were created to automate the avoidance of mismatches in the geocoding process.

C	D	E	F	G
Historical Address	Current Address	Note	X	Y
72 Queen S	72 Queen St S			
20 King W	20 King St W			
76 King St E	150 King St E			
29 Queen S	29 Queen St S			
67 King E	125 King St E			
84 King W	84 King St W			
80 King St E	154 King St E			
9 Foundry S	9 Foundry St S	OANF	-80.49033209	43.45022904
82 King St W	82 King St W			
21 King E	33 King St E	OANF		
31 Courtland W	31 Courtland W			
59 King E	117 King St E			
35 Queen N	35 Queen St N	OANF	-80.48791135	43.45083501
59 King E	117 King St E			
64 Foundry S	120 Ontario St S			
86 King E	86 King St E	OANF	-80.48729521	43.44948724
1 King E	1 King St E	OANF	-80.48870982	43.44969286
1 King W	1 King St E	OANF	-80.48870982	43.44969286
46 King W	46 King St W			
48 King E	70 King St E			
72 King W	72 King St W			
45 King E	45 King St E	OANF	-80.48828413	43.44937765
13 Weber E	13 Weber St E	OANF	-80.48715509	43.45150931
16 Margaret	16 Queen St N			
18 Queen N	18 Queen St N			
221 King W	221 King St E			

Figure 21. Updater csv – with OANF and X,Y

QA Stage 5

QA Stage 1 is repeated on a cursory level and then the Business and Residential csv's are passed on to the next stage of the process, geocoding.

Geocoding

The final component of this project before the data can be imported into the web map is obtaining the latitude and longitude coordinates of the residents and the businesses. Selecting the right geocoder required some research and trial and error. The first iterations of transposed historical addresses were geolocated using *DMTI 2018 Geolocator*, *ESRI Geolocator* and the *QGIS Plugin Geolocator* to see which attained the lowest number of miss-matches. These tests exposed the large number of changes in the street names and address numbers leading to the QA processes and code being created to cleanup and modernize the address data.

First Iteration: ArcPro/ArcGIS Online using shapefiles only

The first iteration to build a proof of concept used the address point shapefile for the City of Kitchener (2019) and spatially joined it to the Teranet parcel fabric polygon shapefile. This was done in an effort to offer a larger area to select (click on) (***Figure 22).***

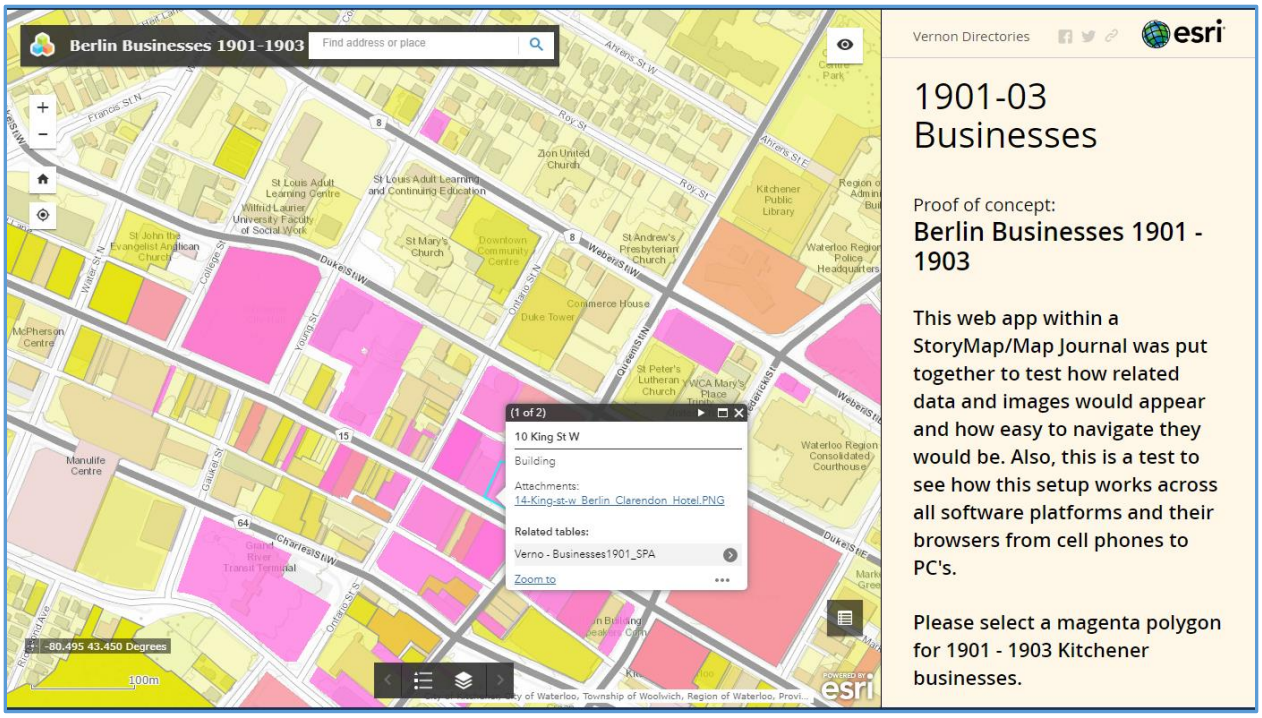


Figure 22. Proof of Concept using ArcGIS Online

Now with the property polygons having an address field the next step was to relate (join) the directory's csv to these polygons. The relate was done as One-to-Many to account for many persons or businesses being at one address. Addresses that didn't match were discarded with no real opportunity for correcting the mismatch. The ESRI products allowed for attachments also to link advertisements to a location on the map.

Unfortunately, relating the shapefile parcel tables to the directory csv proved functional but unwieldy. Familiarity with ESRI products and their layout was a must to understand how the data was presented. As seen in ***Figure 23***, the property polygon is clickable and the address is shown with an attachment available but the other businesses at that address are in the related tables: dropdown, which subsequently needs to clicked to open the information the user is looking for.

Figure 23. Property polygons can only show one table at a time

Geocoding without using an address locator proved functional but not user friendly. Furthermore, re-relating addresses to properties using relational database topology rules would be very time consuming and difficult to automate.

Second Iteration: Address Locator and Shapefiles

The next iteration was to use a local address locator within ArcPro to create X and Y points. Since the team still wanted to use property polygon shapefiles, the DMTI address locator within ArcPro allowed for correcting mismatches and offsetting the points created, off the road center line and onto the property polygons. This worked in theory, but the points created were at times pushed onto the incorrect properties. When the files were spatially joined there was no easy way to know if the points were attached to the correct property. Additionally, storing and presenting hundreds of shapefiles on ArcGIS Online was becoming a cost issue that could be circumvented by not using shapefiles.

Third Iteration: Global Address Locator Software as a Service (SaaS)

Leveraging the use of address locators offers a greater flexibility than the relative strictness of database joins/relations. With the data not being QA'd to the extent that it was later in the project's development, ESRI's address locator was finding more address matches than the local geolocator. This was of course welcomed, however it was using many ESRI credits. Since the addresses needed to be re-geocoded multiple times throughout the QA process, it was using too many credits to be viable.

The team moved on to using the Google / Open Street Maps *MMQGIS Plugin* for QGIS. This proved to be the best of both worlds in that it had less miss-matches and was free to use multiple times. The lower miss-matches were due to two reasons: The QA process was becoming much better for modernizing old addresses, and, that the plugin was placing points at the end of roads instead of reporting miss-matches. These issues will be explained further in the next section as the team looped back to experimenting with the QGIS geocoder more than once.

Fourth Iteration: Address Locator Creating csv's Optimized for Database Storage and Web-based Maps

The Open Street Map *Nominatim Api* service through QGIS was considered as a potential candidate for geocoding the csv files. One issue found with this geocoder is that it would require full addresses (including the street classification St., Ave, Dr., etc.) and cardinal direction, otherwise it would simply match to a general location on a street, or a public transport station. One other issue found when matching using this service is that it will match the street name to some other point of interest's characteristics. An example of this was when trying to match a Water St address, it matched to several public areas that shared a characteristic of water, those being a public splash pad, a pond, a lake, etc. Another issue discovered is that it would add all records, regardless of a match. When no match could be found for the data, the geocoder would cluster the remaining records on a point along the highway, making it difficult to identify unmatched records that required corrections. Finally, the geocoder also had issues with matching to specific addresses, often either tripling the number of records or simply just freezing, depending on whether single or multiple match options were checked. Disappointed with the results, the team moved on to *DMTI 2020 Geocoder* to see if it would meet the project's needs.

DMTI 2020 Geocoder was built using DMTI's large road network which very accurately places points based on the street number given in the address field. The major advantage it holds over *Nominatim* is that it counts the address numbers along a street and does not get held up on locations with a similar name. The DMTI geolocator also has a helpful scoring system when geocoding that helps in distinguishing between good matches, okay matches, bad matches, and no matches. When a record is geocoded the geolocator will give a score from 0 to 100 based on how close the inputs for the address are to the matched address, as well as creates and fills a status field with *M* for matched, or *U* for unmatched. The scoring system is useful for sifting through the records that were mismatched or missing information. Testing has shown that there are approximately 4 different scores that can be given to records that have been put through the geocoding process; 0 which is unmatched, 62.xxx which is matched but the cardinal direction of the matched address differs from the input address, 68.xxx which is matched but the input address is missing the street

type classification (St., Dr., Ave.), and 80.xxx which is successfully matched. These scores are assigned based on a string generated during the geocoding process, which consists of the matched street address, matched city, and two instances of the matched state/province. This string is then compared to the inputs given to the geocoder to see how closely they match. The successful match score will only ever reach a high of 80.xxx due to the generated matched address string containing two state/province sections in the string. The matched address string that is generated by the geocoder is always written as STREET ADDRESS, CITY, STATE/PROVINCE, STATE/PROVINCE. Due to this second state/province, the string will never match 100% of what the user inputs, meaning the best score the process can obtain is 80.xxx. Happy with this geocoder's accuracy, along with a relatively fast time to process and geocode large amounts of data, and with no additional costs, the team decided to move forward with *DMTI 2020 Geolocator*.

The process for geocoding address lists is simple: geocode the list, move the unmatched records to a new file to investigate further (i.e. find closest modern address), re-run through the geocoder, repeat until everything has been matched. It is also necessary to go back and look up the 62.xxx and 68.xxx scored records and correct them. ***Figure 24*** shows the geocoding tool with the fields filled in, and ***Figure 25*** shows an example of what the Matched_Address string looks like.

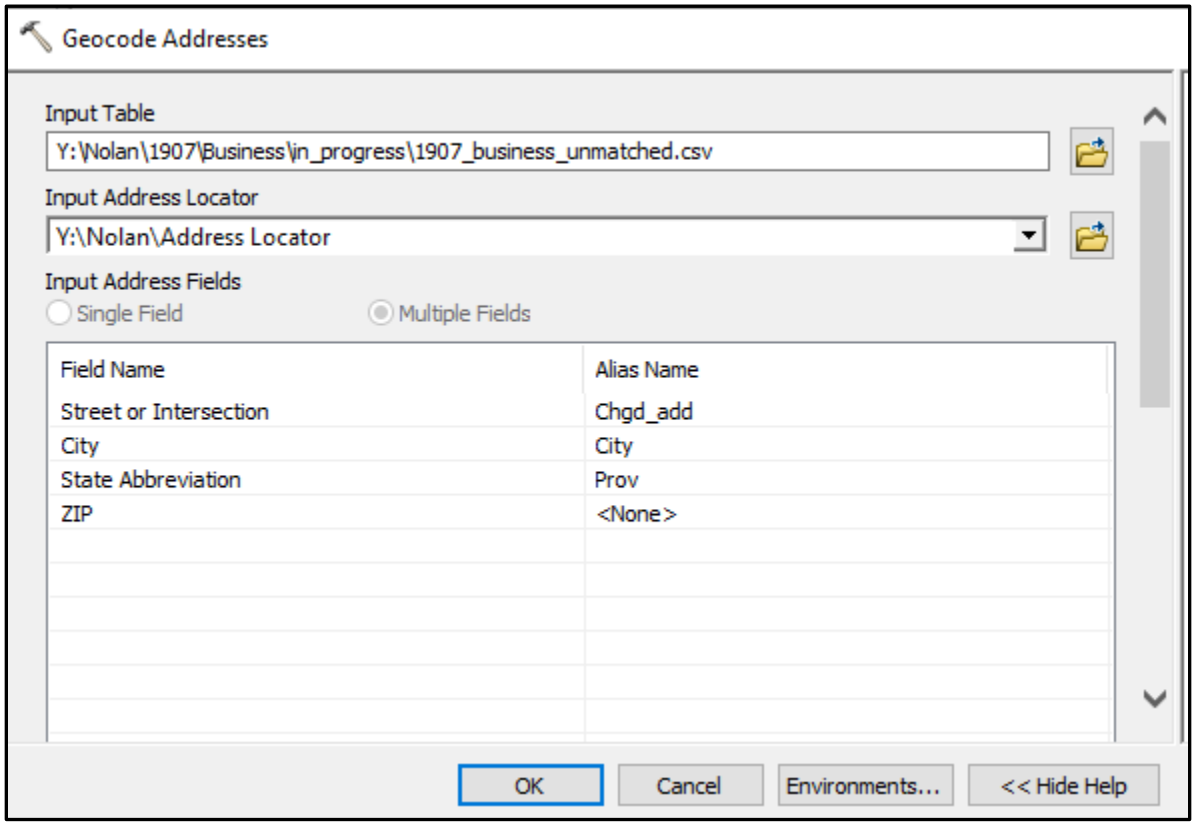


Figure 24. Geocoding tool with the fields filled in

Match_addr
209 FREDERICK ST, KITCHENER, ON, ON

Figure 25. Example of what the generated Matched_Address string looks

Correcting Unmatched Records

Unmatched records are almost always a result of the address no longer existing. The street name and/or number has changed since it once existed and its modern equivalent needs to be determined and replaced. Finding the modern street name requires in-depth research as discussed earlier in this paper. If the street number simply doesn't exist anymore, likely due to demolition and re-development, finding the closest modern address is a simple process, however does require some thought and interpretation, making it impossible to automate.

Addresses in question are searched by entering the address into Google Maps. This will typically place a point on the map that corresponds to where that address should be. Since this is typically done when there is no direct match for the current address, the point will be in one of several places:

- placed in the middle of a street or intersection. This is usually because the street number does not go as high or low as the address being looked up. The address recorded is the one closest to that location (on the same side of the street).
- placed on the side of the road but not on any of the buildings. This is usually when the number should exist in that spot, but the actual numbering skips it. The address recorded is the one that is closest to it.
- placed on the side of the road on open land between buildings. This is because the street number skips the desired address, but there are also no buildings near. This will usually be due to an open field or parking lot. Typically in this situation, the address recorded is the one that occupies that land (i.e. nearby building), or if it is a general parking lot or an open field, then the coordinates are taken directly from Google Maps and entered into the csv (manual geocoding).

One issue that arose while searching for modern equivalent street numbers is that some historical buildings that were demolished were replaced by significantly larger building(s) that occasionally took up the entire block. What was once a building close to a corner, ended up taking up the intersecting street corner as well, and thereby changing the street name altogether (XX Street A to YY Street B). These cases also needed to be dealt with manually.

Lastly, since this project only encompasses the city of Kitchener, any addresses that land in Waterloo need to be removed from the database. This is done through a simple visual analysis of what lands within the limits of Waterloo and deleting those features in the GIS software. The original unmatched data is then marked with a note stating that the original address was not found (OANF), merged with the originally matched data, and exported as a final list. A second list of only the records marked with the OANF note is also compiled and used to update the updater csv, which, as explained earlier, is used to filter tables containing future years of data that has not been worked on yet. This ensures that data from future years will contain more correct current addresses, enabling the initial geocode attempts to yield more matches, requiring less time and resources from team members.

Results and Research Opportunities

As project team members meticulously checked row of data for quality (QA), especially after the discovery of the significant changes to the street address system, they started developing an intimate understanding of Kitchener's geographic and industrial development. By following some addresses and people over time, a lot of information was extracted, many that matched some of the questions that have been asked by library users over time. Although the purpose of this project is to offer research opportunities to others, staff can easily now answer those popular questions like, "what is the history of this building? What was it before it was a grocery store"? "Is this the original building"? "What is the history of my house"? These questions can be answered by following an address throughout all the available directories. ***Figure 26*** is a snapshot of a few businesses over the years. ***Figure 27*** highlights an address with visible changes, courtesty of Google Streetview. ***Figure 28*** is a sample record of those who have resided at one address over the years.

12 King St. E	1897- 1901	W. H. Becker and Co (book seller)
	1907-1912	Beck & Schell Grocers
	1919-1922	Schell Brothers Grocers
	1923-1948	Filsinger & Henry (men's furnishings)
	1949-1964	Hatland Millinery
8-12 King St. E	1965	Andrew Bros Jewellery Ltd.
8 King St. E	1971	Andrew Bros Jewellery Ltd.
	1981-1982	Athlete's Foot
	1985	Vacant
	1990-1994	Knar Jewellery
	2001	Schmid Jewellers
2 King St. W	1897-1901	Sugarman & Co., dry goods
	1907-1908	F. E. Macklin, dry goods
	1910-1911	A J Roos, Druggist
2-8 King St. W	1919-1921	Merchants Bank
	1922Preser	Bank of Montreal
52 King St. W	1897	Pearl Laundry
	1901	H. E. Schreiter, Furniture
	1907-1911	J McGillawee, M D
	After 1919	Address did not exist
45 Queen St. S	1909-1912	Residential
	1919-1920	Customs Exmg Warehouse
45-47 Queen St. S	1921-1925	Customs Exmg Warehouse
	1926	Vacant
	1928-1970	Residential
	1981-1982	No address
45 Queen St. S	1985-2001	Café Mozart

Figure 26. Businesses over the years

Figure 27. Numerous restaurants tried to establish themselves in this building. Prior to housing restaurants, this original building was Canada Trust (1992-1993), Nelco Ltd (plumbing and heating, 1957-1990), Wm Knell & Co (hardware store, 1919-1955), and in 1907 it was Economical Fire Ins. Co.

1909
Harry Koehler, Packer, Rubber Factory
Mannassa Thoman, tmstr

1911
James Blevins, Policeman

1912
J. Smith, Mechanic

1919
Albert Lang

1920-1921
John Warnholz, Kaufman Rubber Company

1922
Hy Caesar

1923 - 1929
Gustave Roemer, Caretaker at St. Peter Church

1930-1931
J. Lloyed (m: Amelia), Mechanice

1932
Murray Steeves

1933
Vacant

1934
Brooke Uttley
Harold Uttley
Jos T. (Selma) Uttley
Lorne Uttley

1935
Edward Sagan, Ontario Shoes
Brooke Uttley
Harold Uttley
Jos T. (Selma) Uttley
Lorne Uttley

1936-1939
Ignatius (Ruth) Young, crvr Baetz Bros Furniture

1940
John (Evelyn) Schuller, Galloway Furniture

1941
John (Ludwina) Schmalz, Proprietor at Schmalz Realty Co.

1942
Ignatius (Ruth) Young, plshr Baetz Bros Furniture

1945
Thos O (Florence) Sparks, A & C Boehmer Ltd.
Evelyn Sparks, Assembler at Marsland Radio

1946
Gladys Knischewski, opr Huck Glove
Thos O (Florence) Sparks, A & C Boehmer Ltd.

1947
Thos O (Delphine) Sparks, Former, Lang Tanning
Thos O (Florence) Sparks, A & C Boehmer Ltd.

1948
Douglas Killian, Norton's Drinx
Thos O (Florence) Sparks, A & C Boehmer Ltd.
Wm K Sparks, Machine Hand at Wunder Furniture

1949
Thos O (Florence) Sparks, A & C Boehmer Ltd.

1950
Thos O (Florence) Sparks, A & C Boehmer Ltd.
Victor W. Sparks, Schneiders.

1951
Chas Sparks, Schneiders
Thos O (Florece) Sparks, A & C Boehmer Ltd.

1952
Chas Sparks, Schneiders
Thos O (Florence) Sparks, A & C Boehmer Ltd.

1953
Thos O (Florence) Sparks, A & C Boehmer Ltd.

1954
Chas (Dorothy) Sparks, Schneiders
Thos O (Florence) Sparks, A & C Boehmer Ltd.

1955 - 1979
Thos O (Florence) Sparks, A & C Boehmer Ltd.

1956
1981-1982
N. Laslavic

1985
R Manning

1990
No return

1994
J Johnston

2001
M. Groff

Figure 28. Sample record of residents over the years at one address

One can of course also follow a specific resident throughout the years. In the example in ***Figure 29***, staff followed Mr. Norton E Staebler, from 1935 to 1948. In those years Norton resided at four different addresses. Three of the four addresses don't exist anymore since the properties were demolished to make room for a parking lot and two apartment buildings, as can be seen using Google StreetView. The latest address on Merner Avenue still shows the original house standing.

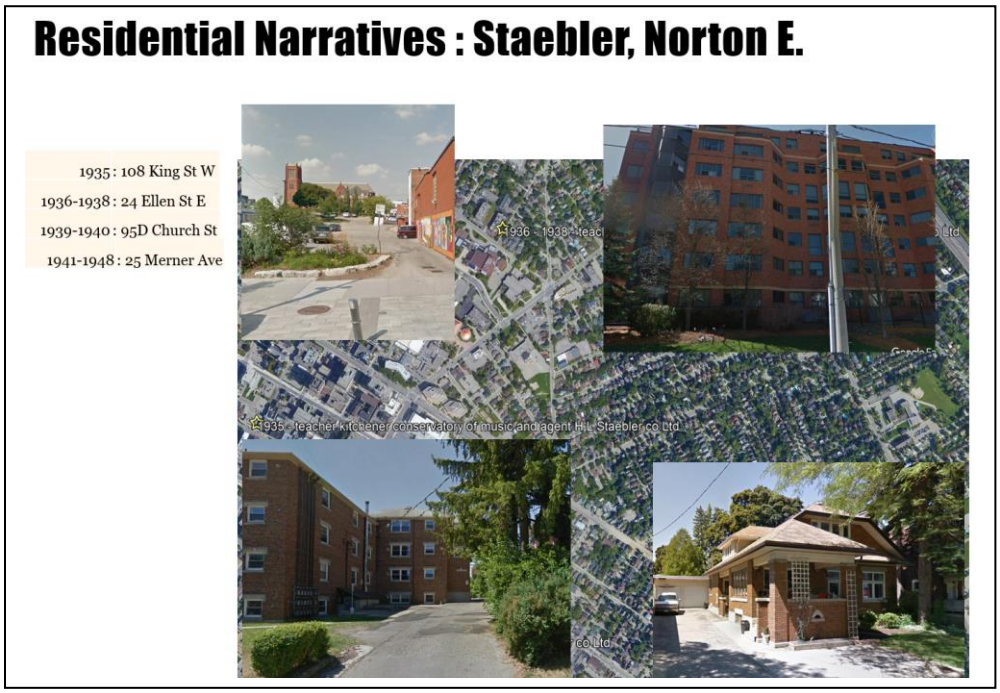


Figure 29. Norton E Staebler's places of residence between 1935-1948

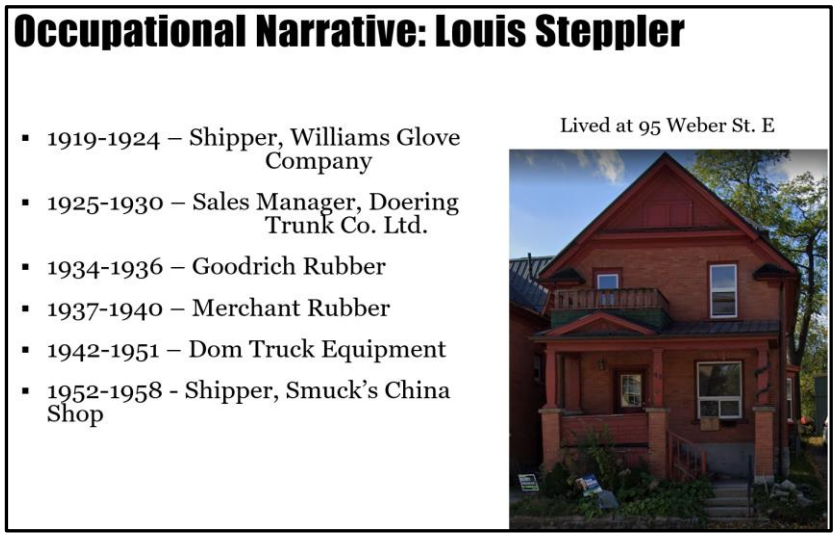


Figure 30. Mr Louis Steppler's career progression

Moving to different addresses was quite comm- on in the early 1900s. Many residents moved every year. This could have been because they frequently changed work. When staff followed residents and their occupations they discover- ed that many did in fact frequently change jobs as well. ***Figure 30*** shows Mr. Louis Steppler's occupations over the years.

When sifting through all the different occupations and places of employment, staff discovered some unique positions. The Rosedale Creamery had a "tester", and confectionary shops had "dippers". The tire factories had "time keepers", and there was always the need for a "sidewalk inspector", "baggage transporter", "peanut vendor", "sidewalk cleaner", "tool sharpener", "basketmaker", and of course "milkman". What was odd to discover however was that occupation progressions didn't always have a natural path. For example, an accountant one year was a piano salesman the next. An engineer became a fireman. A watchman became an electrician. A stenographer turned into a nurse. Perhaps schooling and training was not required as much and therefore allowed for continuous job variety. There were lots of service-type positions, like piano tuners, typewriter and sewing machine repairers, hemmers, and stitchers. Stenographers were predominantly female, as well were dressmakers and operators.

Staff also learned a lot about the manufacturers. Many changed names over the years and some clearly employed a large percentage of the residents in the city. Berlin/Kitchener is probably best known for manufacturing shoes, rubber products, and furniture. Some well known names include Hydro City Shoe, Oscar Rumpel, Dominion Tire, Dominion Rubber, Merchants Rubber, Kaufman Rubber, Canadian Consolidated Felt, Hibner Furniture, Ahrens Shoe. ***Figure 31*** lists and summarizes some of the manufacturers and businesses in 1919 in Kitchener. ***Figure 32*** and ***Figure 33*** show them displayed in a map format.

Clothing	Berlin Robe and Clothing	16-22 Foundry St.
Combs	Berlin Comb and Novelty Co.	Gateman Place
Construction - Concrete til	Shoemaker & Co.	
Electrical	Berlin Electrical Company	
Flour Mill	Shirk & Snider	
Foundry	Berlin Foundry	36 Water St S.
Furniture	Anthes Furniture Co	
Furniture	Berlin Furniture Co.	Victoria St.
Furniture	Berlin Interior Hardwood	72 Wilmot St. W
Furniture	Canadian Furniture Manufacturing	
Furniture	Erb & Co	
Furniture	Hibner Furniture Co.	39 Edward St.
Furniture	Lippert Furniture Co.	222 Louisa St.
Furniture	Kreiner & Co	26 Wilhelm St.
Furniture	Krug Furniture Co. Ltd	111-113 Ahrens St.
Furniture	Baetz Bros & Co.	
Furniture	Simpson Co Ltd	
Furniture	JB Snider	King St. S.
Snyder Bros Upholstering S	Snyder, Roos & Co	
Furniture	Berlin Woodenware Co.	20 Cedar St. N
Furniture	Berlin Table Manufacturing Co.	225 Queen St.
Gas	Berlin Gasoline Engine & Thresher C	209 Queen St. S.
Glass	Cloisonne Glass Co.	62 Foundry St. S
Glove	Waterloo Glove Manufacturing	99 King St. E
Glue	Berlin Glue Works	Strange St. S
Gauntlets, caperines and c	Berlin Gauntlet Co.	16 Frederick St.
Ladders	Berlin Wooden Ware Co.	
Leatherette	Berlin Leatherette Co.	
Mattress	Berlin Mattresss Co.	
Mirror	Berlin Plate Glass and Mirror Co.	30 Foundry St. N
Piano	Foster-Armstrong Co, Pianos	246 King St. W
Plumber and pipe supplies	Forwell Foundry Co.	
Rubber	Berlin Rubber Manufacturing	
Rubber	Kaufman Rubber Co. Ltd	188-198 King St. W
Rubber - boots and shoes	Merchants Rubber Co. Ltd	51 Breithaupt St.
Shirt	Berlin Shirt & Collar Co.	35 King St. E
Shirt	Williams, Green & Rome	
Shirts	Hagen Shirt & Collar Co.	128 Wilmot St.

17 Butchers
9 Button manufacturers
1 carriage maker
6 cleaning and pressing
11 candy and ice cream shops
12 Dentists
11 dress makers
20 furniture makers
13 garages
3 glove manufacturers
Over 140 insurance companies!
20 music teachers
34 nurses
6 restaurants
4 rubber manufacturers
7 shirt and collar makers
15 shoe repairers
12 tobacconists
4 vet surgeons

Figure 31. List of businesses in 1919

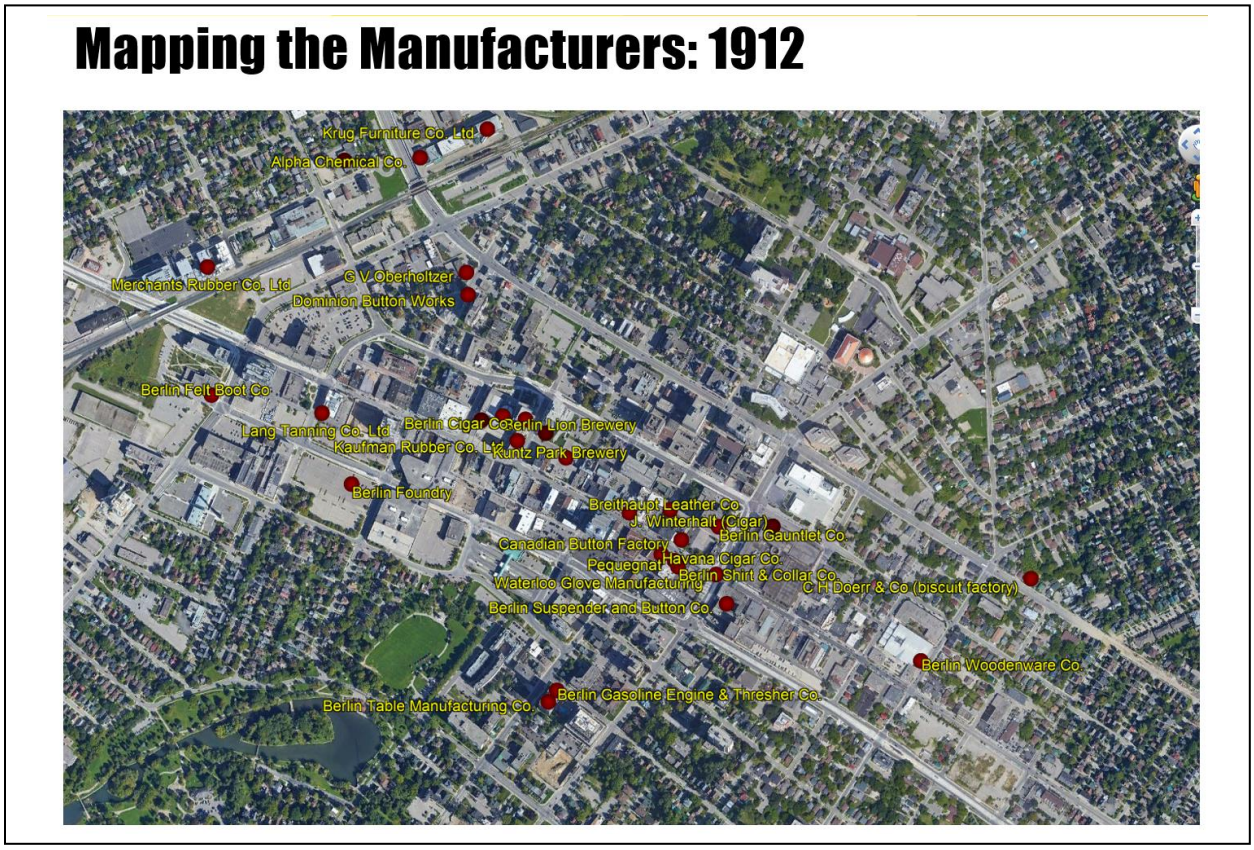


Figure 32. Mapping the Manufacturers

Figure 33. Mapping historical businesses and using Google StreetView to appreciate the buildings are still standing

The city directories can also provide insight into the type of positions available at an organization. ***Figure 34*** shows snippets of people and their positions at the Walper House hotel (still in existence).

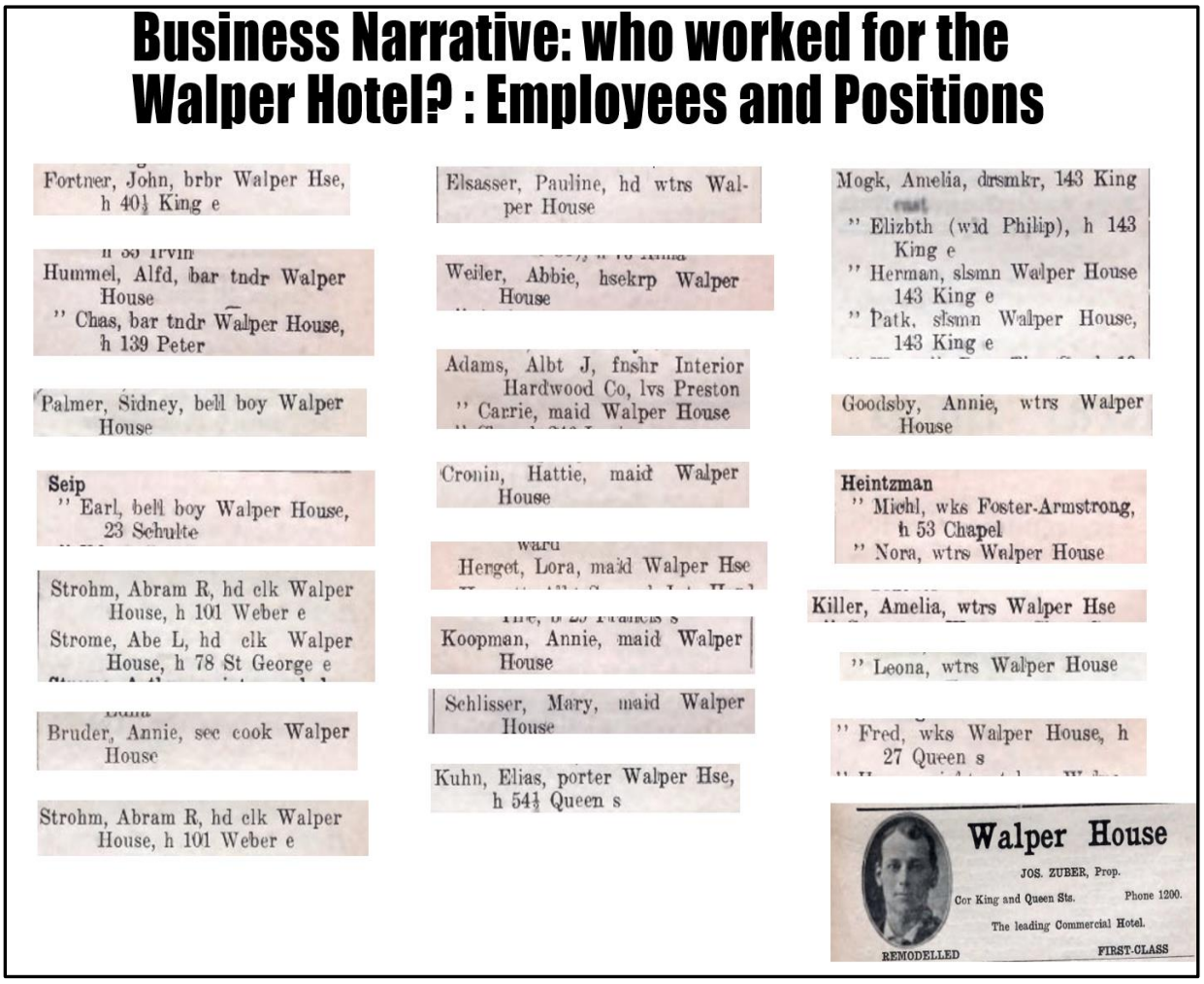


Figure 34. Entries from the early city directories showing employees at the Walper House Hotel

Since these entries were geocoded, one can easily map where these employees live to conduct investigations on home-work travel. ***Figure 35*** shows an example of this. Although the map doesn't show this, it wasn't uncommon for hotel employees to have a room at their hotel, often denoted as a boarder 'b' in the directory.

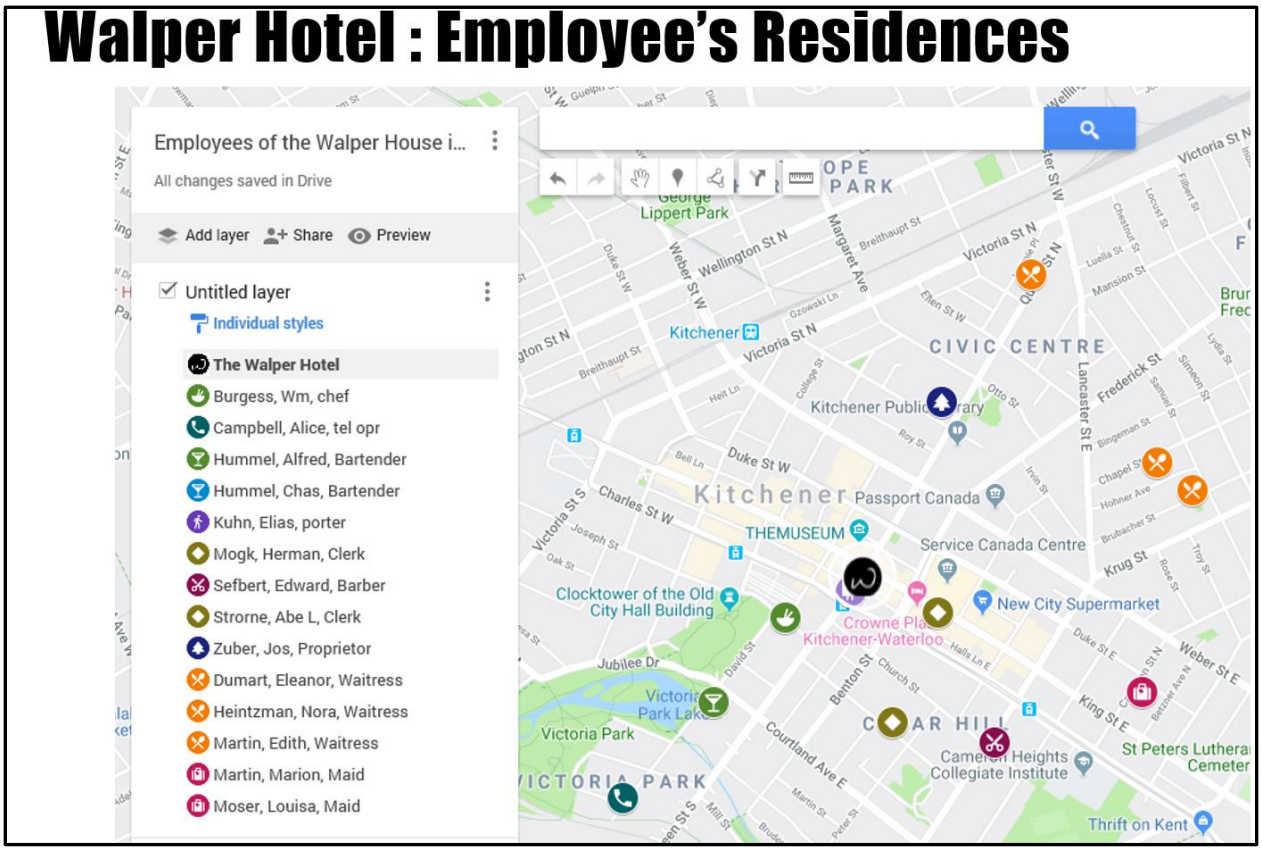


Figure 35. Walper House Hotel employees' places of residents

Advertisements

This modernization of 'old data' not only preserves the past, but in fact opens it up to those who don't normally work with historical print materials. When one thinks of a city directory, they may think of the old white pages or yellow pages, however city directories also include advertisements. These are absolute gems as they reveal types of goods, costs of goods and inflation, business longevity, and style of writing and advertising. The advertisements were all extracted into individual images and have been linked to the business listings. ***Figure 36*** shows examples of some of the advertisements and the types of information that can be extracted from them. ***Figure 37*** focuses on one business, Pearl Laundry, a dry cleaner that was in business for over fifty years. From the ads one can learn that shirts were re-banded for 10 cents in 1901, and that mending was done free, and that decades later the business also offered rug shampooing. ***Figure 38*** shows advertisements that reveal the cost of goods at the time.

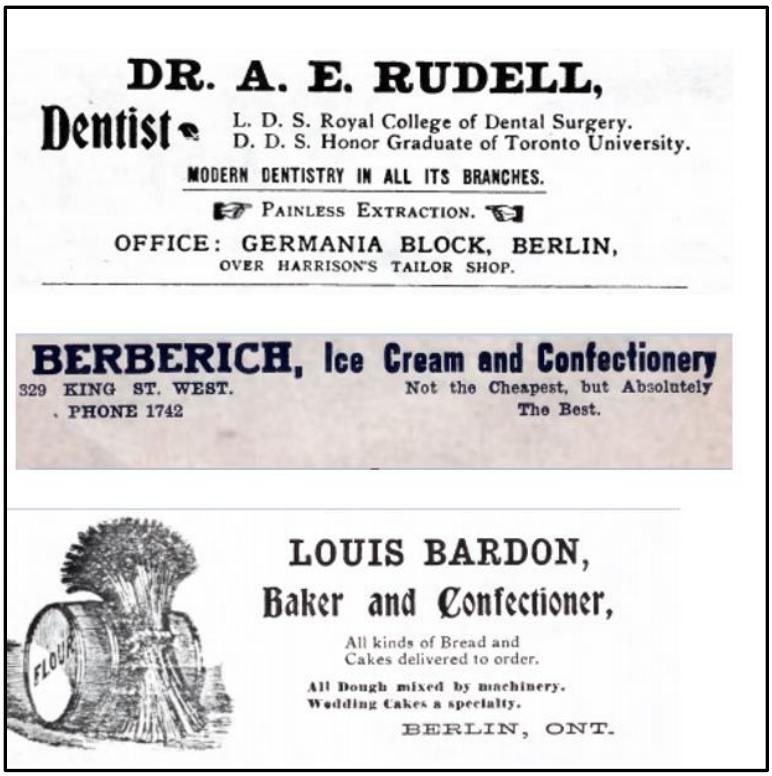


Figure 36. Advertisements from early 1900s city directory. Note some of the descriptions

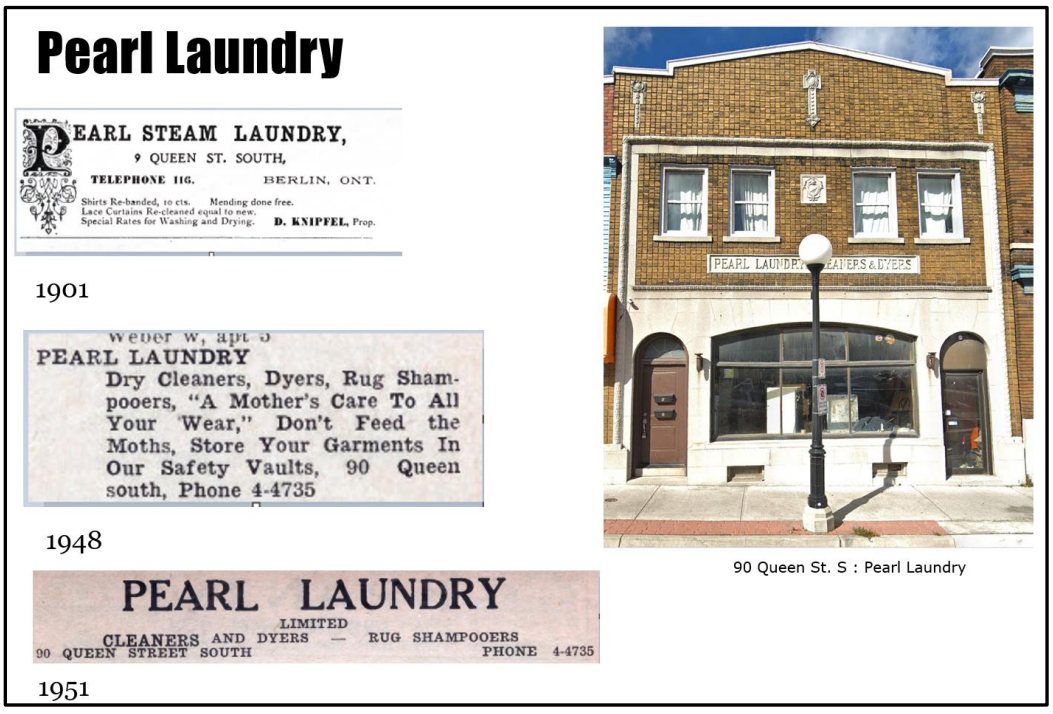
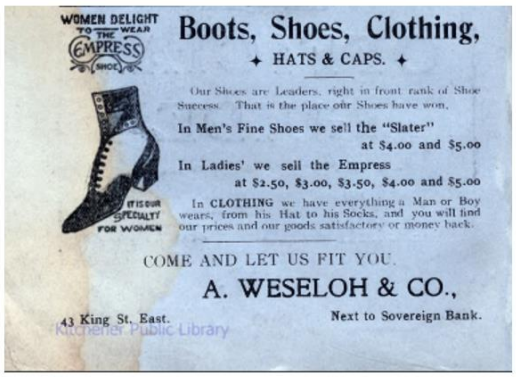
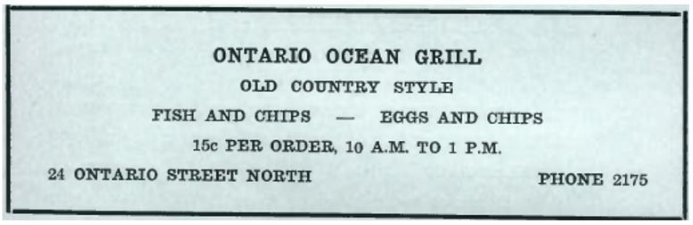


Figure 37. Pearl Laundry advertisements throughout the years



1907

1901



1935

Figure 38. Goods and services capturing prices at the time

Further Research Opportunities

The primary goal of this project is to make rich historical resources easily accessible to the public so that they have the opportunity to ask and answer their own questions about the people and places within Kitchener. Below are just some of the many potential research opportunities that can be explored further using the data extracted from the city directories.

People Narratives

- Types of trades, professions, and services
- Females in the workforce and female entrepreneurship
- Tracking job or career progression
- Journey to work – studying spatial relationships between home and work
- Impacts of war on families – number of soldiers, widows, or address changes
- Studying ethnic groups and their impact on culture
- Studying residential mobility
- Studying prominent family names (and streets named after them)
- Demographics – marriages, re-marriages, deaths
- Urban geography studies – population increase and rates
- Studying urban development and street changes (parking lots, widening of streets, LRT)
- Tracking property re-development and building demolitions
- Studying original structures – an inventory and analysis of historical buildings
- Studying neighbourhood development and street changes

- Personal house research – list of who has lived at a specific address and what they did for a living
- Genealogy – following the surnames of relatives over the years and their migration patterns
- Community snapshots – what were communities like at a specific time – what services did they use? What evidence was there of specific ethnic groups?
- Study of land use planning
- Ecological studies – studying urban sprawl and deforestation

Business Narratives

- Business longevity – A census of stores, businesses and industries. Which businesses were successful? Are there certain areas in the city that had steady long-term commercial establishments? Who were the key employers?
- Industrial development and manufacturing – What did Berlin/Kitchener manufacturer?
- Location of potentially environmentally hazardous sites (i.e. sites of historical gas stations)
- Goods and services – Using ads to gain insight about what was being produced, sold, the cost of goods and tracking inflation
- Studying restaurant establishments – When did the trend of eating outside the home start occurring? Do we have diversity in ethnic restaurants? When did franchises begin?
- Rise of Insurance Companies (based on number of advertisements)

Conclusion

The Berlin/Kitchener and Waterloo city directories provide a rich resource of information about the local communities. Residential information like name, address and occupation is released for every residential address. Advertisements enhance the business information, providing service/product offerings, indicating dates established and listing the proprietors, as well as, in some cases, the executive members. Financial data is provided for insurance companies and banks as well. Translating these data to a map via geocoding could provide some illuminations on the communities and the 1900-1950+ era of Canadian history.

The idea and concept of transcribing and mapping this information is much different than execution mainly due to unexpected hurdles in the street address changes. Additional staff joined the team to help with this, turning this project into a multi-departmental initiative. Dozens of library staff members have been exposed to the concept of geo-location and geocoding and working closely with them has brought working opportunities that otherwise wouldn't be there. This project has also uncovered an important fact – that the city of Kitchener and their residences don't have any paperwork trails of street number changes, as well as many street name changes prior to 1950. A side benefit is being able to offer a street change lookup, through the online project, but also as its own entity. A special online directory will be made available so anybody

interested in learning about the history of streets in Kitchener can correctly do so without having to go through the meticulous examination and investigation that the team members have already completed.

The next step of the project is to complete a few more decades of datasets before releasing the map platform publicly. The methodology for building the map interface (using Leaflet) will be made available at project completion.

The team members look forward to sharing the completed project – a searchable and downloadable interactive online map that organizes the advertisements, household and business entries by street address, enabling query and location-specific searches, allowing researchers to conduct spatial analysis, gain new insights, ask new questions based on geographic proximities, and transform their research and teaching as it relates to urban structure and population geography.

References

Basar, I. (1999). Directory publishing in Canada: the last hundred years. *The Serials Librarian*, *37*(1), 59-82.

Educative (2021). The Levenshtein distance algorithm. https://www.educative.io/edpresso/the-levenshtein-distance-algorithm.

Gregory, I. N., & Healey, R. G. (2007). Historical GIS: structuring, mapping and analysing geographies of the past. *Progress in human geography*, *31*(5), 638-653.

Hardwick, S. W. (1990). Using city directories to teach geography. *Journal of Geography*, *89*(6), 266-271.

M.S. Boehm & Company Ltd (1913). Map of busy Berlin: showing property of M.S. Boehm & Co., Ltd Queen's Park in red. Toronto, Ontario: The Map Company.

Rodger, R., Fleet, C., & Nicol, S. (2010). Visualising urban geographies. *e-Perimetron*, *5*(3), 118-131.

Rose-Redwood, R. S. (2008). Indexing the great ledger of the community: urban house numbering, city directories, and the production of spatial legibility. *Journal of Historical Geography*, *34*(2), 286-310.

Shaw, G. (1982). Nineteenth century directories as sources in Canadian social history. *Archivaria*, *14*, 107-121.

Vernon's Berlin, Waterloo and Bridgeport Directory (1901-1913). Hamilton, Ontario: H. Vernon.

Vernon's Kitchener-Waterloo city directories (1919-1947). Hamilton, Ontario: H. Vernon.

Appendix A

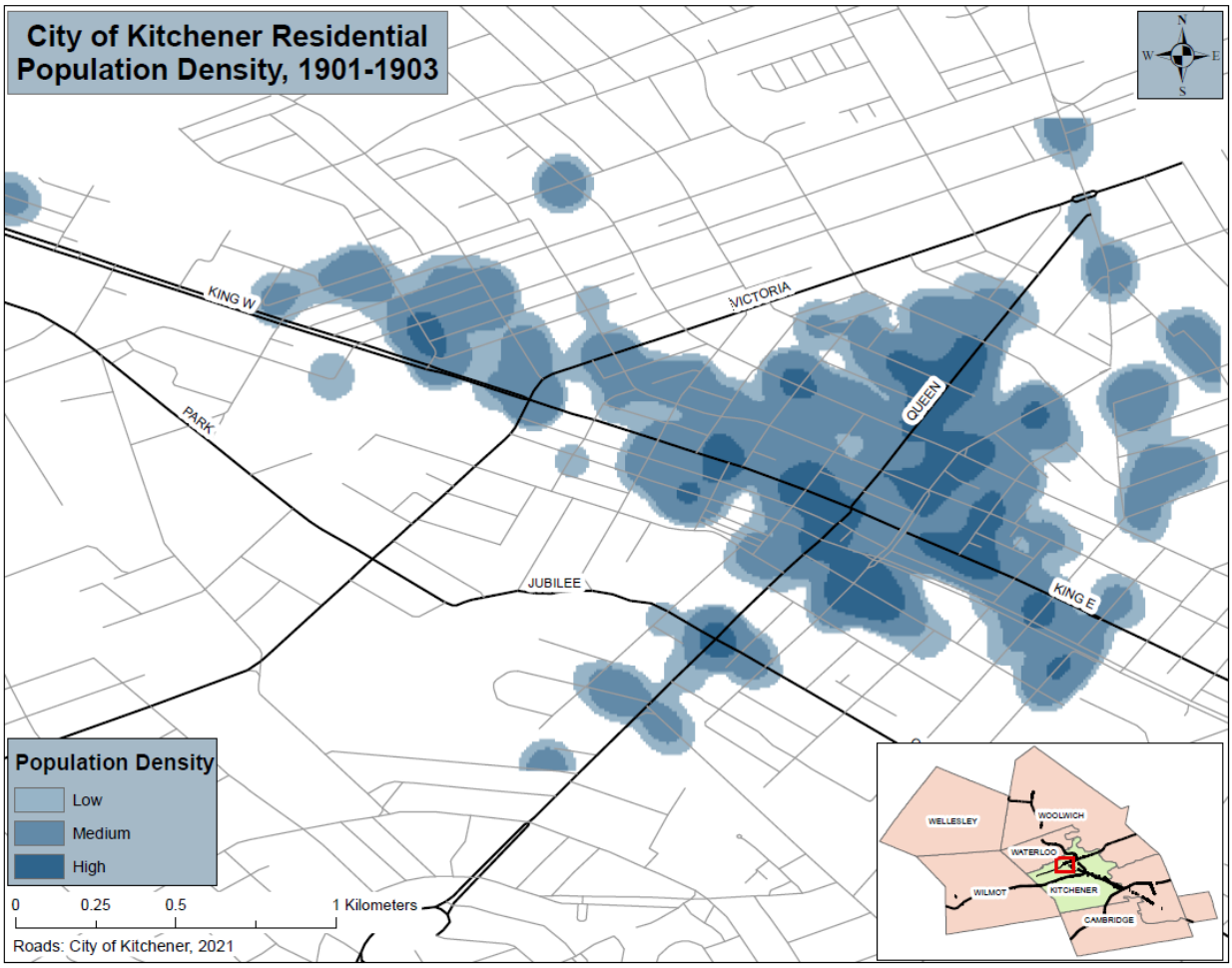


Figure 1. Population density for the years 1901-1903 in Kitchener, Ontario. Many people lived in the downtown area, along King Street and Queen Street

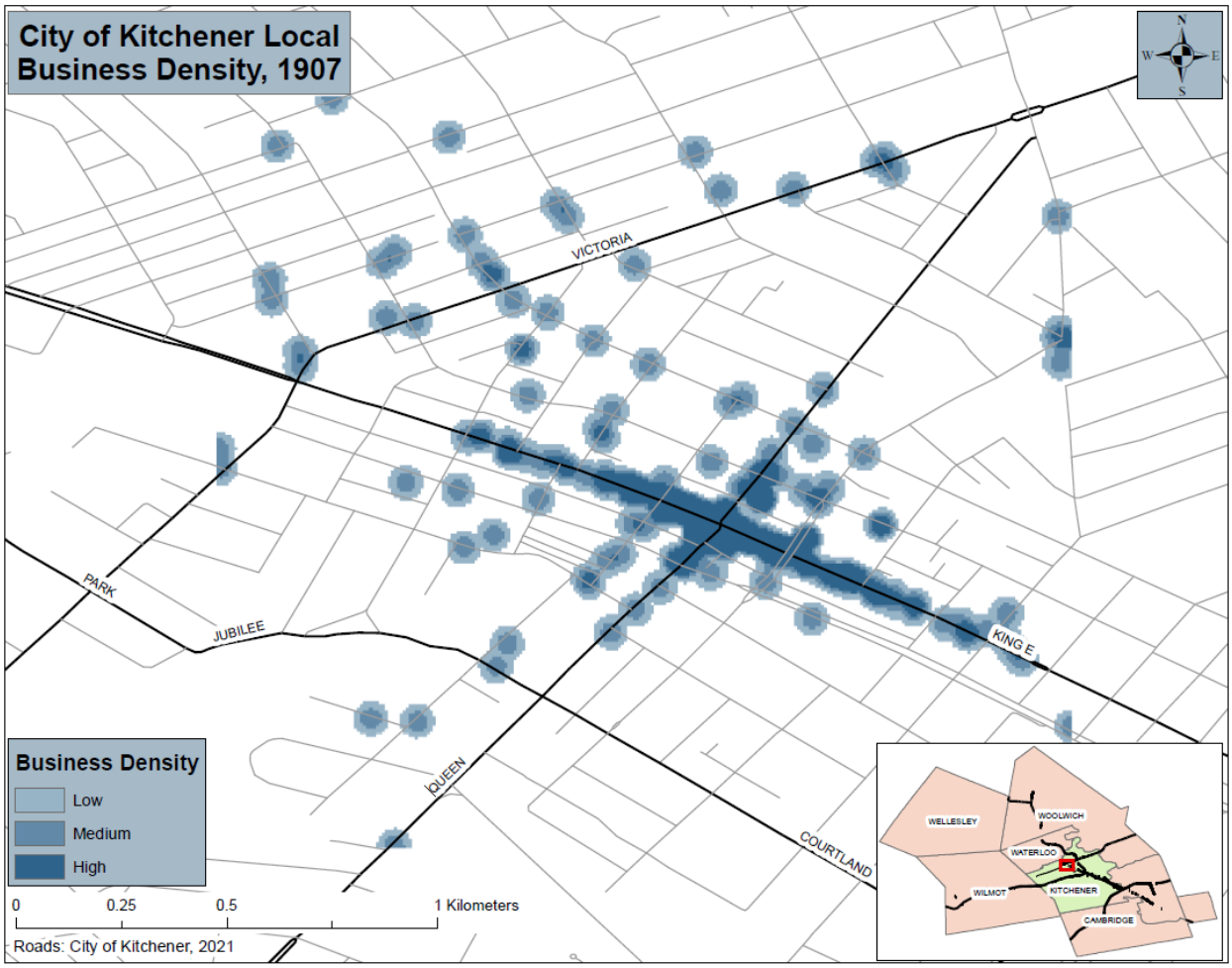


Figure 2. Business density for the years 1901-1903 in Kitchener, Ontario. Businesses were prominently established along King Street, Queen Street and starting to expand out to Victoria Street

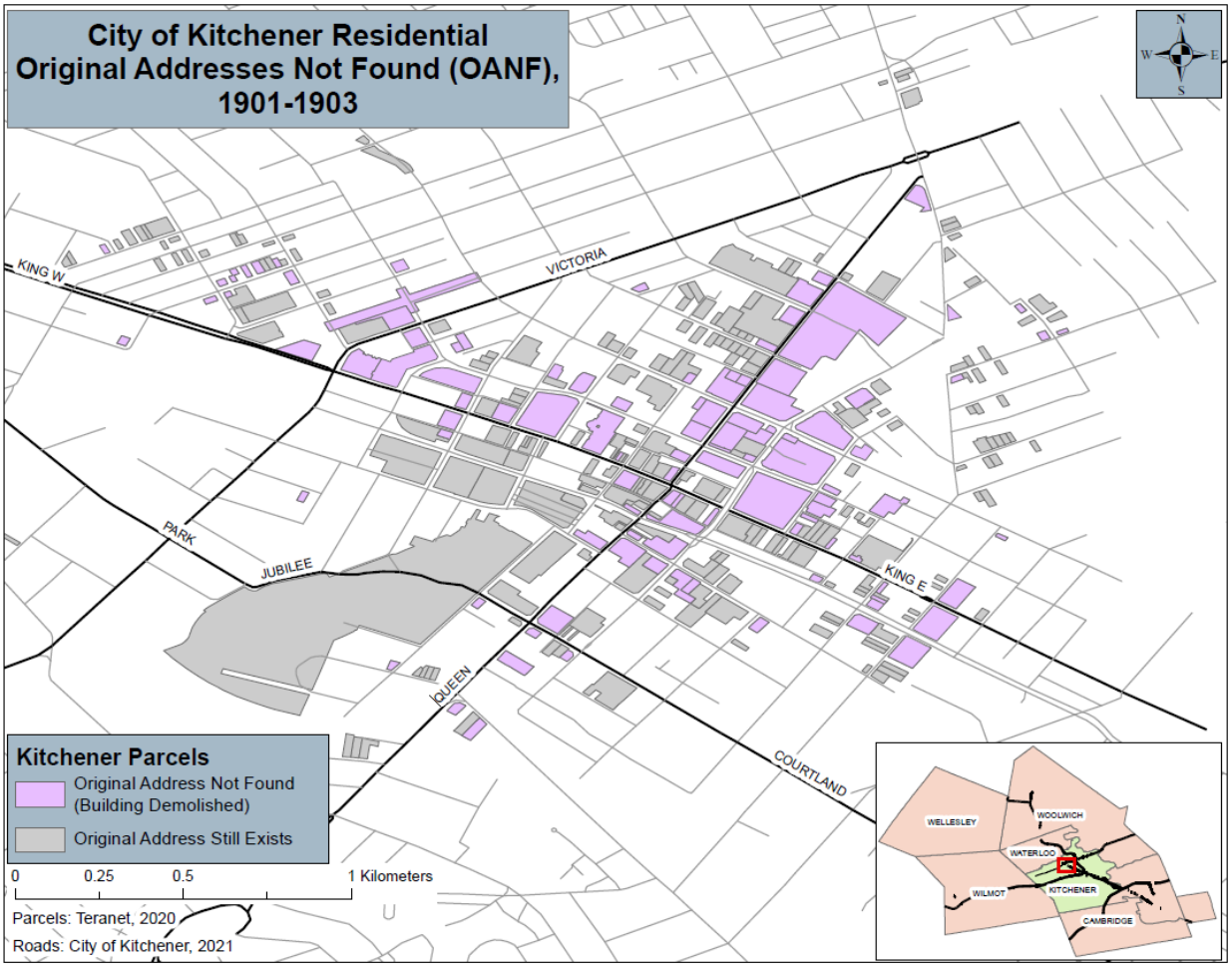


Figure 3. Highlighting the properties that have changed since 1903. Original buildings are no longer present