Erika Whinihan
Data Curation Protocol Report
LIS 545 Winter 2021
March 3, 2021

Municipal Recycling Program Comparison

Protocol Report

1. Project Abstract

This data curation protocol compares the residential (single-family and multifamily) recycling programs for three municipalities, namely Seattle, WA, Portland, OR, and Los Angeles, CA. Recycling is defined as "the process of collecting and processing materials that would otherwise be thrown away and turning them into new products" (Tiseo, 2020). As the climate change crisis continues to gain more urgency, cities can promote and focus on recycling to reduce the strain on the environment and in doing so should provide relevant information and data to the public. These three cities were chosen as a way to compare recycling programs between progressive West Coast cities to provide citizens and residents of these locations a way to see how effective recycling is and how governmental data is stored and made accessible (or not) to the public. As will become evident, recycling data among cities varies a great deal in the way data is tracked and made available to the public. This dataset that pulled information separately for each municipality shows tonnage of recycled materials, cost information, as well as the recycle rates for Seattle (2014, 2016), Los Angeles (2014), and Portland (2016). Seattle data was pulled for two years so a direct comparison could be made to the other two cities, whose more recent recycling data was unavailable in an open data format. The population and area of each city was also included to provide an additional comparison measure and was obtained from the United States Census Bureau.

Recycling data for Seattle was collected through two main sources and was relatively easy to find through open data sources. Seattle is one of the few cities in the country to embrace an open data initiative and the Seattle Public Utilities is one of 36 departments that uses the City's open data portal (Whittington, et al, 2015). The Recycle REPORT (2020) published by Seattle Public Utilities provided the residential citywide yearly average cost and cost/ton for 2014 and 2016. This document also listed the breakdown of tonnage by material and total tonnage for years 2014 and 2016 as well. (Note: the grand total tonnage for 2014 was off by 1 ton when materials were added together so my figure of 83,234 is different than the 83,233 listed in the document). The residential recycling rates for 2014 and 2016 were obtained from the Seattle Public Utilities 2016 Recycling Rate Report (page 4).

Data collection for Portland and Los Angeles was more difficult and onerous and required hand calculations and contact with city employees in an attempt to obtain the relevant data. The City of Portland 2016 data was obtained as an .xlsx file through the Oregon.gov Waste Composition Study webpage. It contained residential recycled tonnage by material although it included additional material categories that Seattle and Los Angeles did not include, so to keep the data consistent only material tonnage was pulled for paper, plastic, glass, metal, and contaminants (hazardous material) for the adjusted total tonnage (21,631).

The recycle rate was calculated by hand from the same source by dividing the total residential tonnage (79,690) by the total tonnage from all substreams of disposed materials (501,056). The total cost of residential recycling in 2016 for Portland and the cost per ton information was not provided so I attempted to obtain this information through email communication with a City of Portland employee. They did provide me with a total cost figure of over $9 million but were not able to show how that figure was obtained and when divided by the total tonnage the cost/ton seemed like an outrageous figure, so I ultimately decided not to include this data. The City of Portland employee confirmed for me that this recycling data is generally not publicly available. They also could not provide a recycling rate for 2016 but did send me a spreadsheet that showed the residential recovery rate for 2018 was 52%.

The recycling data for Los Angeles was obtained from the Residential Waste Stream by Material Type webpage through CalRecycle. The data is from 2014 (Waste Characterization Tool Limits and Background) and can be filtered by county (Los Angeles) and jurisdiction (Los Angeles (countywide)). From here the total residential totals for paper, plastic, glass, metal, and contaminants (hazardous waste) tonnage were taken (874,964) and this number was divided by the total amount of total residential tonnage (2,565,550) to produce the recycle rate. I did submit a request for public records from CalRecycle for residential recycling rates and cost information but received a report from 2014 for commercial recycling only and the data was not sufficiently broken down by municipality. The websites for the City of Los Angeles were difficult to navigate and information on recycling seems to be reported both at the state level and by cities and counties.

Compiling this data into one dataset and making it available to the public through the GitHub repository (user: ewhinihan; repository: municipal_recycling) will allow for citizens to view the data and gain a better sense of the recycling tonnage, cost, and recycling rates for these three cities. The data could be used for educational purposes as well as personal or collective action to mobilize and contact local officials to push for both an increase in transparency with regards to public utilities data. It could also shed light on the recycle/recovery rates for citizens who may be surprised that the rates are not actually higher considering how common recycling is on the West Coast. Finally, it could be used as an example to show citizens how inconsistent the tracking of recycling data is and that to improve transparency, municipalities should embrace open data platforms.

2. Documentation

a. Structured Metadata: Dublin Core

Attribute	Value
abstract	This dataset is a breakdown of sizes (tonnage), cost, and recycling rates for the cities of Seattle, Portland, and Los Angeles for the years 2014 and 2016. The intended audience for this dataset is citizens who would like to be better informed about recycling program effectiveness. This dataset was created for an

	MLIS class (Data Curation) at the iSchool, University of Washington during winter quarter 2021.
accessRights	Public
accrualMethod	Item creation by owner of the collection.
accrualPeriodicity	Annual
alternative	municipal_recyling
audience	Citizens in the general public.
available	This dataset was made public through GitHub repository on February 26, 2021 and will be available going forward.
conformsTo	Open data source in CSV and XLSX formats.
contributor	Sources collected from online state and local municipal government websites for Seattle, Portland, and Los Angeles. Contributions to the resource can be made by anyone as it is publicly available on GitHub. Primary contributor will likely be creator: Erika Whinihan (erikaw9@uw.edu).
created	Created during February through March of 2021.
creator	Erika Whinihan
dateSubmitted	March 3, 2021
description	This dataset contains raw and normalized data in CSV and XLSX formats along with a PDF description of the curation protocol. This dataset was compiled using publicly available websites for the three cities (Seattle, Portland, and Los Angeles).
educationLevel	This data is intended for adult citizens.
format	CSV/XLSX (raw data); XLSX (normalized data)
identifier	https://github.com/ewhinihan/municipal_recycling
issued	Officially published on GitHub on February 26, 2021
language	en-US
modified	No modifications have been made since it was originally published but keeping this value as it may be needed at a later date.
provenance	No changes of ownership as of March 3, 2021 but keeping this value as it may be needed at a later date.
publisher	GitHub: https://github.com/
references	https://www.seattle.gov/Documents/Departments/SPU/Documents/Recycling_Report_2020.pdf http://www.seattle.gov/util/cs/groups/public/@spu/@garbage/documents/webcontent/1_064754.pdf https://www.seattle.gov/Documents/Departments/SPU/Documents/Recycling_Rate_Report_2019.pdf https://www.oregon.gov/deq/mm/Pages/Waste-Composition-Study.aspx https://www.oregon.gov/deq/recycling/Documents/2018MRWGRatesReport.pdf https://www2.calrecycle.ca.gov/WasteCharacterization/StudyDesign https://www2.calrecycle.ca.gov/WasteCharacterization/ResidentialStreams?lg=1019&cy=19 https://data.census.gov/cedsci/ https://data.nal.usda.gov/data-dictionary-blank-template
rights	None; publicly available data on GitHub and through the above-mentioned references.

spatial	Cities of Seattle, WA, Portland, OR, Los Angeles, CA.
subject	Municipal recycling data about the three cities of Seattle, WA, Portland, OR, and Los Angeles, CA.
temporal	2014, 2016
title	Municipal Recycling Program Comparison
type	Dataset contains information about municipal recycling programs.
valid	Dataset is valid for the years in which it describes but could be used for further research in previous or future years.

b. README File and Data Dictionary

The Readme file and data dictionary can be found at the GitHub repository for this dataset and are also included as separate files per the assignment submission.

URL: https://github.com/ewhinihan/municipal_recycling

c. Explanation of Documentation

The Dublin Core Metadata Initiative (DCMI) schema was chosen for this dataset because it is a common metadata schema that many citizens may be familiar with and its website has helpful explanations for how DCMI works and is maintained. The metadata terms are also clearly explained with examples, which made applying it to this dataset manageable and relatively straight forward. The Project Open Data metadata schema would have also been a good choice for this dataset as it is often used for government data and many of the attributes are similar to DCMI.

The metadata attributes that were left out include: bibliographicCitation, coverage, date, dateAccepted, dateCopyrighted, extent, hasFormat, hasPart, hasVersion, instructionalMethod, isFormatOf, isPartOf, isReferencedBy, isReplacedBy, isRequiredBy, isVersionOf, license, mediator, medium, relation, replaces, requires, rightsHolder, source, and tableOfContents. The reason being that not all attributes are required to be included and as the data curator, I have the ability to determine which metadata elements to include to describe this data. The attributes left out also did not seem applicable to this dataset or were redundant in nature.

The Readme file (Readme.md file instead of Readme.txt file as GitHub default is the .md file format) includes a description of the dataset, why it was curated, how data was normalized, as well as the file naming convention. Information about variable names and definitions were left out of the Reame.md file as they are included in the data dictionary, which is provided as a separate file.

3. Data

Please refer to GitHub Repository for data files (raw and normalized), Readme.md file, and data dictionary.

URL: https://github.com/ewhinihan/municipal_recycling

4. Reflection

If I had known how difficult it would be to find three municipalities that tracked recycling data in a consistent way, I likely would have chosen another subject. Compiling this small dataset proved to be quite challenging for many reasons. Comparing the recycle rates for Seattle, Portland, and Los Angeles was hard and it did not seem that the method of calculating by hand provided accurate recycle rates for Portland and Los Angeles. Both these cities show in other documentation (that was not broken down by tonnage so I did not use it) that recycle rates for Portland were as high as 70% for 2015 (Portland Recycles!, 2017) and Los Angeles claims to have recycle rates as high as 76.4% (Recycling, LA Sanitation & Environment, 2021), but again documentation is incomplete and missing in an open data format. The City of Seattle's recycle rates seemed more accurate and realistic as they were provided on the City's website and were listed in a clear way (tables broken down by year from 2000-2016, tonnage, cost, etc.). The process for comparing these values proved to be quite frustrating as I wanted an easy way to provide this percentage as it is a key statistic for citizens to understand the effectiveness of recycling programs across municipalities. As a result, I feel that only Seattle's recycle rate can be fully trusted in my dataset and that the hand calculations will lack long-term viability, sound data provenance, or meaningful reusability.

After spending a great deal of time looking at recycling data, it is clear that most of this information is stored in unstructured and sometimes semi-structured formats. Even the City of Seattle's recycling data, which was the easiest to find and extract information from, exists in multiple formats (PDFs, text boxes, spreadsheets, charts, images, etc.). All three municipalities lack a structured data format for maintaining recycling data and it would be beneficial if upstream data curation is implemented so data curators are involved in the process of data collection, management, and storage from the beginning. The data documentation for the data pulled for each municipality lacked structured metadata, which made it difficult for me as a user to reuse the data with confidence. There was descriptive information for each city in various forms although the sources I pulled municipal data from would benefit from a data dictionary to clarify what values and attributes are provided.

Expansion of the dataset I created to include additional cities and more recent recycling information would be beneficial for citizens to see how recycling rates and cost change over time and how effective recycling is across many municipalities. I chose to use GitHub for my dataset as it is a public repository that anyone can access, changes are documented, and

keeping this repository here shows that this type of data should be freely available to the public. (I also had not used GitHub in this way before so creating an account and my own repository was a very useful and challenging exercise.) A data management plan could be created as well to include what data should be retained and for how long, if any policies around data privacy should be considered, especially since this data includes residential information, and what data storage and preservation of access needs to occur going forward. Overall, I learned a lot through this assignment and feel more confident in my abilities as a data curator and am looking forward to more challenges and learning in Data Curation II next quarter.

References

2016 Recycling Rate Report. (2017, July 1). Seattle Public Utilities. Retrieved on February 22, 2021 from http://www.seattle.gov/util/cs/groups/public/@spu/@garbage/documents/webcontent/1_064754.pdf

Chism-Winfield, Pete. (2021, February 24). City of Portland employee. Personal email communication.

City of Portland 2016. (n.d.). Oregon Solid Waste Characterization and Composition Study. Department of Environmental Quality / Materials Management / Waste Composition Study. Retrieved on February 18, 2021 from https://www.oregon.gov/deq/mm/Pages/Waste-Composition-Study.aspx (Listed here as an .xlsx file)

Data Dictionary – Blank Template. (n.d.). USDA Ag Data Commons. Retrieved on February 20, 2021 from https://data.nal.usda.gov/data-dictionary-blank-template

DCMI Metadata Terms. (n.d.). Dublin Core Metadata Initiative. Retrieved on February 22, 2021 from https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

Desilver, D. (2016, October 7). *Perceptions and realities of recycling vary widely from place to place.* Pew Research Center. https://www.pewresearch.org/fact-tank/2016/10/07/perceptions-and-realities-of-recycling-vary-widely-from-place-to-place/

Portland Recycles! Progress Report and Action Plan. (2017, August). Bureau of Planning and Sustainability. Retrieved on February 25, 2021 from https://www.portland.gov/sites/default/files/2019-08/swr_ptld-recycles-report2017.pdf

Recycle REPORT. Economic Services, Finance Division, Seattle Public Utilities (2020). Retrieved on February 24, 2021 from https://www.seattle.gov/Documents/Departments/SPU/Documents/Recycling_Report_2020.pdf

RECYCLING. (2021). LA Sanitation & Environment (LASAN). Retrieved on February 25, 2021 from https://www.lacitysan.org/san/faces/home/portal/s-lsh-wwd/s-lsh-wwd-s/s-lsh-wwd-s-r;jsessionid=FxPcpqBg5a0n4OPkgLI-I5DN42dvNnCx1k9t6ovmHQ-T_fM19ygC!-1008808110!-365958312?_adf.ctrl-state=1ddbz1vt1o_1&_afrLoop=3005988898975498&_afrWindowMode=0&_afrWindowId=null#!%40%40%3F_afrWindowId%3Dnull%26_afrLoop%3D3005988898975498%26_afrWindowMode%3D0%26_adf.ctrl-state%3D1ddbz1vt1o_5

Residential Waste Stream by Material Type. (2019). CalRecycle. Retrieved on February 18, 2021 from https://www2.calrecycle.ca.gov/WasteCharacterization/ResidentialStreams?lg=1019&cy=19

Tiseo, I. (2020, October 30). *Recycling in the United States – Statistics & Facts*. Statista. https://www.statista.com/topics/1275/recycling-in-the-united-states/

United States Census Bureau (n.d.). Retrieved on February 25, 2021 from https://data.census.gov/cedsci/

Waste Characterization Tool Limits and Background (2019). CalRecycle. Retrieved on February 25, 2021 from https://www2.calrecycle.ca.gov/WasteCharacterization/StudyDesign

Whittington, J., Calo, R., Simon, M., Woo, J., Young, M., & Schmiedeskamp, P. (2015). Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government. *Berkeley Technology Law Journal*, 30(3), 1899-1966. DOI: 10.15779/Z38PZ61