

Curation Protocol Report

1. Project Abstract

This dataset consists of total recycling tonnage, total refuse tonnage, and waste diversion rates in three U.S. cities: New York City, NY; Providence, RI; and Franklin, TN. The data were searched for and found utilizing the Open Data Network by Socrata, a search engine that allows users to search across open data catalogs, including open data portals for U.S. cities and states. “Waste diversion” was the primary search term used to discover the datasets from New York City, Providence, and Franklin. Population data was pulled from the U.S. Census Bureau and supplemented to enrich the dataset.

The datasets are separately meant to track each city’s waste diversion progress over time. Although the datasets lack sufficient documentation and do not overtly define waste diversion within the context of their individual programs, we can understand waste diversion rates generally to be the amount of waste diverted from incineration or landfill (CleanRiver). Waste diversion rates are critical benchmark data for understanding recycling program processes and performance metrics.

There were many challenges associated with finding and comparing sources for waste diversion rate data across U.S. cities. For instance, it is unclear how cities vary in their method for calculating diversion rates. New York City calculates diversion rates by dividing total recycling by total waste. It seems this formula should be standard across U.S. cities, but a lack of data documentation from other cities obscures how diversion rates are in fact calculated. In general, there is an overarching lack of standardization across these recycling programs. The datasets consist of remarkably different variables, thus making them difficult to parse out and compare. I will elaborate on these challenges in section 4 of this report.

The intended audience for this dataset is investigative journalists interested in reporting on recycling programs in New York City, Providence and Franklin specifically. This dataset could also be reused and expanded upon by journalists interested in collecting and comparing data extracted from the recycling programs of other U.S. cities. In order for journalists to meaningfully use this data, additional research must be conducted to understand how each city calculates diversion rates, what each city considers a recyclable material; etc.

2. Documentation

a. Structured Metadata: Project Open Data

Attribute	Value
title	City Recycling Programs
description	This dataset consists of total recycling tonnage, total refuse tonnage, and waste diversion rates in three US cities: New York City, NY; Providence, RI; and Franklin, TN. It was created as part of an assignment for LIS545 at the University of Washington iSchool.
keyword	"Recycling" "diversion" "municipal waste" "refuse" "New York City" "Providence" "Franklin"
references	https://data.providenceri.gov/Sustainability/Annual-Waste-Diversion-and-Recycling-Rates/mr3j-eu43/data https://performance.franklintn.gov/Government/Diversion-Rate/8rq2-z966 https://dsny.cityofnewyork.us/wp-content/uploads/2019/10/about_d_sny-curb-side-collections-FY2019.pdf
modified	03/08/2021
fn	Tania María Ríos Marrero
hasEmail	mailto:tmrios@uw.edu
identifier	https://github.com/tmrios/city_recycling_programs
accessURL	https://github.com/tmrios/city_recycling_programs/blob/main/RecyclingPrograms_Normal_VS1.csv
downloadURL	https://github.com/tmrios/city_recycling_programs/blob/main/RecyclingPrograms_Normal_VS1.csv
format	CSV
accessLevel	Public
rights	These data are freely available to the public through the open data portals of the respective cities and states.
publisher	Tania María Ríos Marrero

b. README File and Data Dictionary

Please refer to my Github repository for the README file and data dictionary:

https://github.com/tmrrios/city_recycling_programs

c. Explanation of Documentation

The DCAT-US Schema v1.1 (Project Open Data Metadata Schema) was chosen for this dataset because of its common usage by government agencies. I used the Metadata Field Guidance Catalog to explore the required attributes as well as attributes that could be applied to expand and specify the dataset. I included all attributes required by the schema, with the exception of “bureauCode” and “programCode,” which are only applicable if a federal agency is publishing the dataset. Although the “references” attribute is not required in this scheme, I included it because I thought that any person who hopes to use or reuse this dataset would benefit from seeing the primary sources.

3. Data

Please refer to my Github repository for data files:

https://github.com/tmrrios/city_recycling_programs

4. Reflection

As previously stated in the abstract, there were many challenges associated with finding and comparing sources for waste diversion rate data across U.S. cities. I observed an overarching lack of standardization across recycling programs. Datasets consisted of remarkably different variables, and the process of comparing them and identifying matching and/or comparable variables was slow and painstaking.

In general, I was surprised by the limited data documentation provided by New York City, Providence and Franklin. Of the three, New York City provided the most documentation, but it was still sparse in places. New York City calculates diversion rates by dividing total recycling by total waste. It seems this formula should be standard across U.S. cities, but a lack of data documentation from other cities obscures how diversion rates are calculated.

Similarly, there was a lack of documentation from each city regarding which materials are considered recyclable and counted towards the total recycling tonnage. For example, from looking at the original dataset, one can infer that New York City recycles metal, glass, and plastic (MPG) as well as paper, but it is unclear whether they include organics (leaves, yard waste, food waste) in their recycling count. In order for journalists to meaningfully reuse the dataset I curated, there is a need for additional research to understand what materials each city recycles, as well as what each city considers to be

non-recyclable materials. A greater understanding of these variables will enhance our comprehension of values for total recycling tonnage and values total refuse tonnage.

One of the most challenging things I came across was the fact that New York City calculated it's recycling and refuse tonnage daily, whereas Providence and Franklin reported annually. This difference posed a challenge to effectively combining these datasets. As data curators, how do we account for this important distinction when designing and formatting our datasets? I considered creating labels stating "per day" vs "annually" but could not decide on the best approach, so left as is and simply included the distinction in the README file. I am left wondering how to best treat this.

Overall, this assignment has led me to a greater understanding and appreciation for the complexity and value of data curation. I am also astounded by the lack of standardization and reporting across municipal recycling programs in the U.S. It is clear that the curatorial work of aggregating and making data reusable and shareable is critical to the progress of government programs and strategic plans.

Sources

Annual Report: New York City Curbside and Containerized Municipal Refuse and Recycling Statistics by Borough and District. (2019). New York City Department of Sanitation. Retrieved 6. Mar. 2021 from https://dsny.cityofnewyork.us/wp-content/uploads/2019/10/about_dsny-curbside-collections-FY2019.pdf

Annual Waste Diversion and Recycling Rates. City of Providence Open Data Portal. Retrieved 6. Mar. 2021 from <https://data.providenceri.gov/Sustainability/Annual-Waste-Diversion-and-Recycling-Rates/mr3j-eu43/data>

Diversion Rate. City of Franklin, Tennessee. Retrieved 6. Mar. 2021 from <https://performance.franklintn.gov/Government/Diversion-Rate/8rq2-z966>.

- - - - -

"Best Practices for File Naming." *Stanford Libraries*, library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming. Retrieved 6. Mar. 2021.

DCAT-US Schema v1.1 (Project Open Data Metadata Schema), 6 Nov. 2014, resources.data.gov/resources/dcat-us/.

"What Is A Waste Diversion Rate And Why Is It Important?" CleanRiver, 15 Mar. 2019, cleanriver.com/blogwhat-waste-diversion-important/. Retrieved 8. Mar. 2021.